2009

Stereotype Threat and Implicit Attitudes: Implications for the Leaky Pipeline of Women in Science

Karen E. Stamm
University of Rhode Island

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Stamm, Karen E., "Stereotype Threat and Implicit Attitudes: Implications for the Leaky Pipeline of Women in Science" (2009). *Open Access Dissertations*. Paper 918.
https://digitalcommons.uri.edu/oa_diss/918

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

STEREOTYPE THREAT AND IMPLICIT ATTITUDES:
IMPLICATIONS FOR THE LEAKY PIPELINE OF WOMEN IN SCIENCE

BY

KAREN E. STAMM

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN
PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2009

ABSTRACT

Historically, women have been under-represented in science disciplines. This phenomenon has been described as a “leaky pipeline” where fewer women advance to higher levels of the pipeline. A possible mechanism behind the leaky pipeline is stereotype threat, the fear of being negatively evaluated because of one’s group membership. Stereotype threat can undermine the performance of the members of any disadvantaged group. For example, women may not perform well on science tasks because of the stereotype that “women can’t do science.”

This project examined whether stereotype threat negatively influences performance on an Implicit Association Test (IAT), a test that measures unconscious attitudes. The sample included 311 male and female students and faculty. In order to activate stereotype threat, participants in the experimental condition were asked to indicate their gender and were given instructions that the test is intended to measure scientific thinking. A series of two-way ANOVAs included gender, discipline (science, social science, non-science), or student-faculty status (undergraduate students, graduate students and faculty) as the first independent variable; experimental condition was the second independent variable. Performance on the IAT was the dependent variable, where high scores represented greater implicit endorsement of traditional gender-science stereotypes. Results indicated no differences by gender. The discipline analyses revealed an interaction between discipline and experimental condition. Individuals in the no-stereotype-threat condition (control) from science disciplines showed greater endorsement of traditional gender stereotypes, followed by

individuals in social science disciplines, and individuals in non-science disciplines. There were no differences in the stereotype-threat condition. The student-faculty analyses found that there were differences between undergraduate students and all other participants. Graduate students and faculty in the no-stereotype-threat condition showed greater endorsement of traditional gender stereotypes than undergraduate students in the no-stereotype-threat condition, while there were no differences in the stereotype-threat condition. A structural equation model found that, for men, a latent construct measuring General Academics was a predictor of a latent construct measuring Science Identification, which in turn was a predictor of the performance on the IAT. The structural equation model did not fit as well for women. Implications for women in science and the leaky pipeline are discussed.

ACKNOWLEDGEMENTS

There are many people, organizations, and entities who I wish to acknowledge and thank for their assistance with this project.

I thank the Undergraduate and Graduate Research Grant Committee and the Graduate Assistants United union for funding a significant portion of my research.

I thank my committee for their input and suggestions: Ellen Flannery-Schroeder, Faye Boudreaux-Bartels, John Boulmetis, and Tom Malloy. I thank Tom for encouraging me to turn my class project from his social psychology seminar into actual research. I thank Faye for her helpful comments during the pilot phase of my study. I thank my major professor, Lisa Harlow, for her support and encouragement, not only in this project but also throughout my graduate career. She taught me statistics and how to teach. I consider it a great compliment when a friend told me I sounded like Lisa.

I thank the ADVANCE Program for supporting me as a research assistant and for providing me with the basic idea for my dissertation.

I thank the creators of www.reducingstereotypethreat.org for maintaining a well-organized and up-to-date website that significantly cut down on the time I needed to search (and re-search) the literature on stereotype threat.

I thank my family, in particular my husband Miguel, my parents, and my in-laws, for their ongoing encouragement and support.

I thank running for teaching me everything I needed to know about graduate school. One of the best convergence of events was finalizing my dissertation proposal

while training for a half marathon. The processes were remarkably similar. I successfully completed the half marathon and proposed my dissertation in the same week, both in just under two hours. I repeated the process while I was preparing to defend my dissertation. Again, I finished the half marathon and the defense in under two hours. Graduate school is a marathon, not a sprint. Set goals. Work hard. Persist in the face of pain. Try again if you don't get it right the first time. Celebrate when you meet your goals.

I thank the Halftime Program and the Intersections Project at Boston College for providing me with a framework of three key questions: What are my talents and gifts? What brings me joy? Who needs me to do it? These questions focus on talent, joy, and need. Another way to phrase the questions is: Where have you been? Where are you now? Where are you going? These questions are aimed at identifying vocational discernment, but I have always thought the questions are useful for making a decision at any crossroad. I have returned to the questions many times, and they have been very helpful in helping me to find my way. The most recent set of questions became: What can I learn? What can I do? What do I want to be able to do?

I owe a final thank you to the countless people who helped to put me on the right path. Like most graduate students, I doubted whether I could actually reach my goals. I realized that I picked the wrong path, slowed down, turned myself around, and completely changed my research interests. It was incredibly challenging to switch gears, but it was tremendously worthwhile. I ended up studying exactly what I thought I was going to study when I first came to graduate school: gender stereotypes. I am leaving graduate school as a quantitative social psychologist.

DEDICATION

To justice, self-control, and jumping over the hurdles

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
DEDICATION	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER 1	1
STATEMENT OF THE PROBLEM	1
CHAPTER 2	3
JUSTIFICATION FOR THE PROJECT	3
<i>The Leaky Pipeline: Losing Women in Science</i>	3
<i>Considerations of Social Context</i>	6
CHAPTER 3	9
THEORETICAL FRAMEWORK	9
<i>Social Identity Theory</i>	9
<i>Cognitive Processes: Identity Threats</i>	12
<i>Empirical Research on Stereotype Threat</i>	13
<i>Stereotype Threat and Identification</i>	15
<i>Limitations in Previous Empirical Research on Stereotype Threat</i>	17
<i>Implicit Attitudes</i>	17
<i>Hypotheses</i>	19

CHAPTER 4	23
METHOD.....	23
<i>Participants</i>	23
<i>Measures</i>	24
<i>Procedures</i>	26
CHAPTER 5	30
RESULTS	30
<i>Preliminary Analyses</i>	30
<i>Analyses</i>	31
<i>Gender</i>	33
<i>Discipline</i>	34
<i>Student-Faculty Status</i>	36
<i>Revised Discipline and Student-Faculty Status Analyses</i>	37
<i>Prediction Models</i>	40
<i>Familiarity with IATs</i>	54
CHAPTER 6	56
DISCUSSION	56
<i>Hypothesis 1</i>	56
<i>Hypothesis 2</i>	59
<i>Hypothesis 3</i>	60
<i>Hypothesis 4</i>	61
<i>Comments on the Experimental Manipulation and Other IAT Issues</i>	65
<i>Stereotype Threat and Women in Science</i>	66

<i>Implications for the Leaky Pipeline</i>	70
<i>Limitations</i>	73
<i>Future Directions</i>	76
<i>Conclusions</i>	78
APPENDICES	110
Appendix A. IAT Instructions for Blocks 1-7	110
Appendix B. Categories and Items Used in the IAT.....	112
Appendix C. Measures	113
Appendix D. STEM and Non-STEM Disciplines.....	116
Appendix E. Informed Consent Form	117
Appendix F. Recruitment Flier	119
BIBLIOGRAPHY	120

LIST OF TABLES

TABLE	PAGE
Table 1. Selection of Empirical Literature on Stereotype Threat in Women in Science	81
Table 2. Implicit Association Test (IAT) Block Overview.....	82
Table 3. Procedures and Experimental Design	83
Table 4. Frequency Crosstabulations for Gender, Discipline, and Student-Faculty Status and Chi-Square Results	84
Table 5. D Score Means and Standard Deviations for Gender, Student-Faculty Status by Experimental Condition	85
Table 6. Two-way ANOVA Source Table for Gender, Discipline, and Student-Faculty Status by Experimental Condition Analyses.....	86
Table 7. Frequency Crosstabulations for Gender, Revised Discipline, and Revised Student-Faculty Status and Chi-Square Results.....	87
Table 8. D Score Means and Standard Deviations for Revised Discipline and Revised Student-Faculty Status by Experimental Condition.....	88
Table 9. Two-way ANOVA Source Table for Revised Discipline and Revised Student-Faculty Status by Experimental Condition Analyses	89
Table 10. Correlation Matrix for Identification Survey Items and D Score for the Full Sample	90
Table 11. Correlation Matrix for Identification Survey Items and D Score for Men ..	91

Table 12. Correlation Matrix for Identification Survey Items and D Score for Women	92
Table 13. Means, Standard Deviations, and Reliability Estimates for Identification Survey Items	93
Table 14. Hierarchical Multiple Regression Results for Men	94
Table 15. Hierarchical Multiple Regression Results for Women	95

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Theoretical Model of Under-representation of Women in STEM Disciplines.....	96
Figure 2. Diagram of Anticipated Prediction Model	97
Figure 3. Means Plot for Gender and Experimental Condition	98
Figure 4. Means Plot for Discipline and Experimental Condition.....	99
Figure 5. Means Plot for Student-Faculty Status and Experimental Condition.....	100
Figure 6. Revised Means Plot for Discipline and Experimental Condition.....	101
Figure 7. Revised Means Plot for Student-Faculty Status and Experimental Condition.....	102
Figure 8. Mediational Model with Measurement and Structural Paths for the Full Sample.....	103
Figure 9. Mediational Model with Measurement and Structural Paths for Men	104
Figure 10. Full Model with Measurement and Structural Paths for Men	105
Figure 11. Direct Effects Model with Measurement and Structural Paths for Men ..	106
Figure 12. Post-Hoc Adjusted Mediational Model with Measurement and Structural Paths for Men	107
Figure 13. Mediational Model with Measurement and Structural Paths for Women	108
Figure 14. Post-Hoc Adjusted Mediational Model with Measurement and Structural Paths for Women.....	109

CHAPTER 1

STATEMENT OF THE PROBLEM

Diversity is important in the pursuit of scientific knowledge (National Science Foundation, 2005a). Historically, women have been underrepresented in the sciences. Although women's participation in the sciences has increased, advancement has been uneven. Fewer and fewer women continue to higher levels of professional advancement, a phenomenon that has been described as a "leaky pipeline" (National Science Foundation, 2005b).

The theoretical framework behind the leaky pipeline of women in science is rooted in social identity theory (Tajfel, 1981). Social identity theory explains some of the mechanisms related to the functioning of the social self and the need for positive social identities (Schmader, 2002). The theory examines the social self in relation to group membership, identification, social comparison, and self-esteem needs. Identity threats, such as social identity threat and stereotype threat, extend from the theoretical framework and can undermine participation in the sciences.

Social identity theory lays the foundation for stereotype threat, which is the idea that performance can be impaired because of a negative stereotype regarding a group to which an individual belongs (Steele, 1997). One way to study stereotype threat is through the use of implicit associations, which are attitudes that exist outside of awareness (Nosek, Banaji, & Greenwald, 2002a). Both stereotype threat and implicit associations can operate outside of conscious awareness. However,

stereotype threat and implicit associations have a key difference. Stereotype threat only affects the members of the disadvantaged group. For example, women may be impaired in science tasks because of the stereotype that women do not do well in science domains. Stereotype threat would not have the same effect on men attempting a science task. In contrast, implicit associations are held by both stereotyped and non-stereotyped group members. For example, men and women are equally likely to endorse the stereotype that women do not perform well in science domains.

This project examined whether performance on an implicit association task is negatively influenced by stereotype threat. That is, individuals who experienced stereotype threat were expected to be negatively influenced and show implicit associations consistent with traditional stereotypes. Additionally, the project examined whether stereotype threat plays a role in the leaky pipeline of women in science. It is important to measure the effects of stereotype threat through indirect measures (Greenwald et al., 2002), such as implicit associations, in order to prevent response bias that can be found in explicit attitudes. An experimental manipulation was intended to activate stereotype threat in women in science, thereby leading to longer reaction times on an implicit association task. A prediction model was also examined to understand factors associated with science identification. It was anticipated that the results would illuminate a possible mechanism behind the leaky pipeline, specifically the stereotype that women do not perform well in science.

CHAPTER 2

JUSTIFICATION FOR THE PROJECT

The advancement of science requires a variety of viewpoints. It appears that the gender gap in the sciences is narrowing (Feist, 2006). However, at all transition points in career advancement, a smaller proportion of women move from lower levels to higher ones. This phenomenon, a gradual loss of women, is termed the “leaky pipeline” (National Science Foundation, 2005b). Pipeline issues become apparent beginning in middle school and extend throughout the highest levels of professional advancement, such as membership in national academies of science. For example, in engineering, the percentage of bachelor’s degrees earned by women increased steadily until the late 1990s but has stalled recently at 20% (Commission on Professionals in Science and Technology, 2007). This statistic suggests problems with the pipeline of future scientists. A possible mechanism for the leaky pipeline is the relationship between stereotype threat and implicit attitudes. Threats to one’s identity and unconscious beliefs may undermine women scientists as these women attempt to climb the ladder of career advancement.

The Leaky Pipeline: Losing Women in Science

The leaky pipeline refers to the higher proportion of men that is found at various career levels. Although progress is being made slowly, the underrepresentation of women in the sciences has been a long-standing issue (Pfafflin,

1984). Pfafflin acknowledges the important contributions that women can make to science and technology, such as bringing different perspectives to the table.

Stereotype threat has been observed in educational settings, including group work in middle school students (Huguet & Régner, 2007) and high school classrooms (Keller & Dauenheimer, 2003). Beginning around high school, the leaky pipeline becomes apparent. However, even before high school, gender differences in math achievement have been noted in elementary school students as early as first grade (Penner & Paret, 2008). Small differences can translate into big losses at higher points in the science pipeline. Although young women and men take approximately the same number of science classes in high school, women tend to take more classes in biological sciences and men take more classes in physical sciences (Feist, 2006). In high school, interest differences begin to emerge. Female high school students were more likely to report interest in becoming physicians or biologists, while male high school students were more likely to report interest in becoming engineers or physicists (Lee, 1998). The gender differences become noticeable as fewer women choose science majors as undergraduates. Fewer women pursue graduate studies in the sciences; women receive 39% of masters and 33% of Ph.D.s in the sciences and engineering (Feist, 2006). Finally, fewer women receive tenure-track positions. Women make up approximately 25% of the total science and engineering workforce and hold less than 21% of faculty positions in science and engineering (National Science Foundation, 2005a). Although more Ph.D. degrees are earned by women in science and mathematics than in the past, those advances do not translate into more women at higher academic ranks (National Science Foundation, 2005b; Noordenbos,

2002; Ceci, Williams, & Barnett, 2009). The pattern also occurs in fields that are female-oriented, such as psychology (Kite et al., 2001). For example, men in the social sciences are more likely to be tenured 6 to 10 years after receiving their Ph.D.s than women (Rudd et al., 2008). In academic leadership positions, women are underrepresented in comparison to the number of men who hold leadership positions and to the number of women in the pool for such positions (Dominici, Fried, & Zeger, 2009). The gap is widest at high-ranking professional levels. For example, membership in the National Academy of Sciences is overwhelmingly male (Feist, 2006). There is a need to understand the mechanisms behind the leaky pipeline as well as actions that can be taken to repair the science pipeline.

Many factors may contribute to the leaky pipeline. For example, support from mentors is necessary to assist new faculty in learning the institutional culture (Rice, Sorcinelli, & Austin, 2000). In particular, female mentors in the sciences serve as role models (Marx & Roman, 2002). If possible, multiple mentors can provide a wide range of support. Bias avoidance (Drago et al., 2006) may also play a role in the leaky pipeline. Examples of bias avoidance include planning a pregnancy during the summer so that maternity leave will not be an issue and giving false reasons for backing out of commitments (leaving for “another meeting” to be able to watch a child’s sporting event). The relatively small number of women in lower ranks, hostile campus climate, bias resulting in discrimination, and a lack of attention to the work-life balance are also factors that may result in the leaky pipeline (National Science Foundation, 2005b; Rosser & Taylor, 2009), especially in full professor and leadership positions.

An underlying mechanism behind many of these factors appears to be an implicit attitude that women and science do not go together. For example, a hostile climate sends the unconscious message that women are not welcome in the sciences. Inaccurate perceptions of women may lead to negative evaluations and slow advancement (Valian, 1998). Bias may be subtle and can occur in the form of stereotype threat (Steele, James, & Barnett, 2002), a factor that can undermine performance. Pervasive bias towards women may indirectly contribute to the leaky pipeline of women in science.

Considerations of Social Context

It is important to briefly consider the larger social context in relation to mechanisms behind the under-representation of women in the sciences. Implicit bias plays a role in the larger social context. For example, research on implicit associations between math and gender found that, in a sample of undergraduate students, men were identified with math and women were not identified with math (Nosek, Banaji, & Greenwald, 2002b). The implicit attitude that women and math do not go together sends a message of “Women need not apply.” Similar implicit associations are found in racial attitudes (Greenwald, McGhee, & Schwartz, 1998). Pervasive bias discourages under-represented group members from participating in STEM disciplines.

Prejudice and discrimination also play a role in the larger social context. The promising news is that overt gender discrimination is declining and is increasingly less acceptable. However, subtle discrimination still exists, including benevolent sexism in

which women are portrayed positively but also in need of men's assistance (Dardenne, Dumont, & Bollier, 2007). A study on women's success at stereotypically male tasks found that negative evaluations occurred when women were successful in male-dominated domains (Heilman, Wallen, Fuchs, & Tamkins, 2004). This study suggests that there is an evaluation bias against women, particularly in areas that are not stereotypically female. Women are penalized for violating gender norms. Subtle racism can also work through implicit bias (Dovidio, Gaertner, Kawakami, & Hodson, 2002). Subtle bias can occur in areas, such as hiring choices and college admission decisions. Although the biases may be small, they can accumulate into larger effects (Valian, 1998).

Finally, the surrounding culture influences the social self. Stereotypes make assumptions about what group members ought to do. For example, gender stereotypes generally suggest that women are not good at math (i.e., "Math is hard!" Barbie). Classic work on sex differences (Maccoby & Jacklin, 1974) found gender differences as well as similarities across many domains; these researchers attempted to debunk some of the myths regarding gender differences. Recent work has also shown that most outcomes are characterized by gender similarities rather than gender differences (Hyde, 2005). Still, there are some key gender differences in interest. In a review of gender differences in interest in science, Feist (2006) suggested that differences in interest reflect a difference in the environmental factors that influence gender stereotypes about appropriate disciplines to pursue. Men are more likely to pursue physical science disciplines, while women are more likely to pursue social science disciplines. Stereotypes can indicate the cultural appropriateness of certain disciplines.

The purpose of the current project is to examine the role of stereotype threat in the under-representation of women in science. Several variables, including gender, discipline, and student-faculty status are considered. Student-faculty status is considered because it is expected that there are differences in implicit associations in individuals who are at different places along the pipeline. The goal of the current project is to examine the connection between stereotype threat and implicit associations related to women in science. Additionally, this project will seek to understand the factors related to science identification as well as possible gender differences in science identification.

CHAPTER 3

THEORETICAL FRAMEWORK

The theoretical framework for this project draws on literature on social identity theory, cognitive threats in the form of social identity threat and stereotype threat, and implicit attitudes. Social identity theory (Tajfel, 1981) addresses the importance of group membership to the identity of the individual, while social identity threat and stereotype threat are cognitive processes that can undermine the individual's performance and thus undermine identification with the group. Additionally, stereotype threat can be activated without conscious awareness. That is, an individual does not need to believe the stereotype in order to be influenced by the stereotype (Steele, 1997). Implicit associations also exist outside of conscious awareness (Nosek et al., 2002a). Through the ability to detect subtle forms of bias, implicit associations offer a mechanism to help understand both stereotype threat and social identity theory. Because stereotype threat can occur outside of awareness, a task that measures implicit associations is well-suited to study this phenomenon.

Social Identity Theory

The basic principle behind social identity theory (SIT) is that identity and positive self-esteem are drawn from group membership. The theory states that self-esteem is related to a positive evaluation of one's own group in comparison to other groups (Tajfel, 1981). Social identity is defined as "the part of one's self-concept that

derives from group membership” (Fiske, 2004, p. 438). The major constructs of the theory include social categorization, social comparison, and mechanisms behind positive self-esteem. Social categorization, the classification of stimuli into ingroups (groups to which one belongs) and outgroups (groups to which one does not belong), is used in order to simplify the world (Hogg & Abrams, 1988). Social categorization can rely on stereotypes. While stereotyping is based on some truth, stereotypes are often prone to errors or exaggerations. Social comparison refers to the tendency to look to others or other groups as a means to assess oneself. SIT emphasizes the importance of self-esteem needs. Positive group membership allows for self-enhancement (Fiske, 2004; Schmader, 2002). Also, positive self-esteem is derived from strong group identification, while threats to group identification can lead to a loss of self-esteem (Dietz-Uhler & Murrell, 1998). By promoting positive self-esteem, group membership also promotes positive social identity. There is a strong motivation to maintain positive social identities.

In addition to focusing on the function of groups in meeting self-esteem needs, SIT provides a basis for exploring intergroup relations. Social categorizations are involved in intergroup prejudice and status differentials. Categorization is an automatic process that helps individuals to make sense of the vast amount of information in the environment. Social categorizations both reduce and create intergroup bias (Dovidio, Gaertner, & Validzic, 1998). The roots of prejudice and discrimination may lie in the favoring of one’s in-group. The theory also accounts for differences in status between groups. SIT rests on the assumption that “society comprises social categories which stand in power and status relations to one another”

(Hogg & Abrams, 1988, p. 14). When dealing with two or more groups, a social hierarchy exists. One group has high status, and the other group has low status.

SIT provides a social-cognitive basis for the group categorization and identification processes underlying the under-representation of women and minorities in STEM disciplines. Figure 1 shows an integrated model building on the theory. The social self is the starting point in the model. The social self includes any number of social identities, group memberships, self-categorizations, and roles. Two pathways extend from the social self; identity threats and the influence of the social context are mechanisms that factor into the individual's response. The more direct pathway focuses on social-cognitive threats that are related to social identity threat related to group membership or stereotype threat related identification with a stereotyped domain. Once activated, the threats impair performance (i.e., lower scores on math tests for women who identify with the domain of math), lead to disidentification with the stereotyped domain, and activate mechanisms to protect self-esteem. The indirect pathway focuses on the larger social context, which influences the individual through implicit bias, prejudice, discrimination, and the surrounding culture. Like identity threats, the social context can impair performance, lead to disidentification, and activate self-esteem protection mechanisms. A consideration of both SIT and stereotype threat suggests that, when faced with stereotype threat, fewer women will remain in science domains. The final social outcome is the under-representation of women in the sciences.

Cognitive Processes: Identity Threats

Two types of threats stem from social identity theory. First, social identity threat is a belief that an individual may be negatively evaluated based on the individual's social identity. It can occur under a broad set of conditioning, including when the group's value is undermined (Branscombe, Ellemers, Spears, & Doosje, 1999). The threat does not need to be intrinsic to the individual. Situational cues, such as observing a gender-unbalanced environment (i.e., a social interaction containing more men than women), can activate social identity threat (Murphy, Steele, & Gross, 2007).

Second, stereotype threat is a subtype of social identity threat. When a negative stereotype exists, "anything one does or any of one's features that conform to [the stereotype] make the stereotype more plausible as a self-categorization in the eyes of others, and perhaps even in one's own eyes" (Steele & Aronson, 1995, p. 797). That is, stereotype threat is a fear of being evaluated by or conforming to a negative stereotype about one's group. Steele and Aronson (1995) first demonstrated the effects of stereotype threat. When a difficult verbal task was presented as a test of intellectual ability (a manipulation of racial stereotypes), Black participants performed worse in the stereotype threat condition and performed at a comparable level as Whites in the non-stereotype threat condition. Additionally, the salience of race was increased by asking participants to indicate their race before completing a task. The task was not presented as a measure of intellectual ability. The researchers found that the "mere cognitive availability of the racial stereotype is enough to depress Black participants' intellectual performance" (Steele & Aronson, 1995, p. 808). This result points to the

activation of stereotypes by increasing salience. Even without having a test presented as a measure of intellectual ability, stereotype threat can occur.

Empirical Research on Stereotype Threat

Stereotype threat originated in the study of academic performance in African Americans. Simply increasing the salience (i.e., the relevance to the self) of race created poor performance on a standardized test for African Americans (Steele & Aronson, 1995). That is, having individuals become acutely aware of their distinguishing features (e.g., their gender or race), may instigate stereotype threat, thus leading to a threat to social identity and poorer performance. Stereotype threat has been studied in many other domains, including women in math and science (Nosek et al., 2002a; Nosek et al., 2002b; Murphy, Steele, & Gross, 2007; Kiefer & Sekaquaptewa, 2007; and Kawakami, Steele, Cifa, Phills, & Dovidio, 2008) and a preference for young over old (Nosek et al., 2002a). Stereotype threat is not confined to members of minority groups. It can also occur in White athletes (Stone, Lynch, Sjomeling, & Darley, 1999) and White men when they are compared to Asian men (Aronson et al., 1999). A meta-analysis conducted by Nguyen and Ryan (2008) found that, overall, stereotype threat resulted in performance impairments.

Although activating a stereotype can undermine performance, increasing the belief in an individual's ability can reduce the effects of stereotype threat (Stangor, Carr, & Kiang, 1998). Increasing individual's beliefs in ability resulted in better performance. However, this increase did not occur when group stereotypes were activated. In this case, stereotype threat diminished the individual's perceived ability

and performance. A recent study (Smith, Sansone, & White, 2007) examined the role of interest and motivation in performance. Women who were interested and motivated in science were more likely to avoid science activities when the stereotype was activated. The finding from the Smith et al (2007) study suggests that stereotype threat has a negative effect on performance because it undermines beliefs in competence.

Other recent studies have addressed the role of situational cues, such as the presence of math and science textbooks and the number of men and women present in social interactions. Murphy et al. (2007) examined the threat presented by gender-related situational cues and the salience of group membership, such as identification with the domains of math, science, or engineering. Murphy and colleagues found strong support for the role of situational cues in activating stereotype threat. Women who strongly identified with science domains were affected the most by the situational cues. Men were unaffected by situational cues. Similarly, Kiefer and Sekaquaptewa (2007) found that situational cues have an effect on stereotype threat activation in women. However, situational cues interact with implicit gender-math associations such that women who hold implicit gender-math stereotypes are more likely to be affected by stereotype threat.

Finally, another recent study (Kawakami et al., 2008) investigated stereotype threat in women and behavior related to approaching or avoiding math. Participants who received “Approach Math” training were instructed to pull a joystick toward themselves when shown a math-related image and push the joystick away from themselves when shown an arts-related image, while participants who received “Avoid Math” training were instructed to pull the joystick toward themselves when shown an

arts-related image and push the joystick away from themselves when shown a math-related image. Results indicated that women trained to approach math held more positive implicit attitudes towards math and were more highly identified with math, but this result was only true for women who did not initially identify with math.

Stereotype Threat and Identification

Steele (1997) argues that stereotype threat applies to any disadvantaged group. In light of Steele's research suggesting a link between stereotype threat and gender, it is possible that asking an individual for his or her gender will bring gender stereotypes to mind. By doing so, stereotype threat may affect the individual's performance on activities that are non-stereotypical for his or her gender. As previously discussed, gender stereotypes in the sciences generally hold that men are better at sciences than women (Feist, 2006). This has been demonstrated in both explicit attitudes (Hyde, Fennema, Ryan, Frost, & Hopp, 1990) and implicit attitudes (Nosek et al., 2002b). Solo status, such as being the only member of a minority group, can have an additive effect when combined with stereotype threat in negatively affecting performance in a stereotyped domain (Sekaquaptewa & Thompson, 2003). The double threat of being a lone minority or the only woman and fears about one's performance is a consideration for under-representation in the sciences.

The literature suggests that there is a relationship between domain identification and stereotype threat. Specifically, individuals with greater identification have an increased risk of performance impairments due to stereotype threat. Stereotype threat is highest for those who strongly identify with the stereotyped

domain (Steele, 1997). For example, an individual who strongly believes that he or she possesses math ability is more likely to have performance on a behavioral task negatively affected by a stereotype. For individuals whose identity is closely related to the social group of interest, stereotype threat can more severely impair a behavioral task. The social self comes under attack because of stereotype threat (Marx & Stapel, 2006), especially when the negative stereotype is important to one's identity. Wheeler and Petty (2001) suggest that activation of a self-stereotype is more likely to cause negative performance through stereotype threat. The focus on self-stereotypes in increasing stereotype threat points to the role of self-identity. An intervention to minimize the effect of stereotype threat was effective in creating an identity-safe environment (Davies et al., 2005). That is, stereotype threat did not have an effect when one's identity was not under attack, even if negative group stereotypes were activated. This finding suggests that identity may be an important mechanism for the operation of stereotype threat.

A critical point is disidentification with the stereotyped domain (Steele, 1997). For example, a woman who highly identifies with math is more likely to turn away from the domain if her math performance is undermined by the threat. It is possible that disidentification may occur as a protection of self-esteem. Burkley and Blanton (2008) suggest that self-esteem can be "saved" by endorsing a negative stereotype. Although this conclusion is counterintuitive, the negative stereotype provides an explanation for failure at a stereotyped task.

Limitations in Previous Empirical Research on Stereotype Threat

Table 1 summarizes recent empirical work on stereotype threat in women in science. The table highlights some of the limitations of previous empirical work. The limitations include the sample used. For example, some studies only included women (Kiefer & Sekaquaptewa, 2007; Smith, Sansone, & White, 2007). All of the studies used samples of undergraduate students. Another limitation is that the stereotype threat manipulations varied. For example, one study manipulated situational cues (Murphy et al., 2007), while others manipulated gender priming (Marx & Stapel, 2006; Smith et al., 2007). A third limitation is the definition of math and science. Some studies focused specifically on math (Kiefer & Sekaquaptewa, 2007; Kawakami et al., 2008), while others focused more generally on science (Murphy et al., 2007; Smith et al., 2007). Rather than studying gender stereotypes about math or science, Davies, Spencer and Steele (2005) researched gender differences on a leadership task. Another limitation is that all of the studies were conducted in laboratory environments; none of the studies listed in the table were conducted in a real-world setting.

Implicit Attitudes

Implicit attitudes are unconscious associations, often made regarding categorization (Nosek et al., 2002a). For example, gender stereotypes easily come to mind, even without awareness. Implicit attitudes can have an effect on behavior. In looking at interracial relations, Dovidio, Kawakami, and Gaertner (2002) found that implicit attitudes are involved in behavior that people do not try to control (i.e.,

nonverbal cues). Greenwald and Banaji (1995) call for indirect measures of associations. Since implicit attitudes can be outside of conscious awareness, it makes sense to examine them indirectly. The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a categorization task that measures unconscious associations between two broad categories (i.e., gender and career), each with two levels (i.e., male and female for gender, career and home for the career category). Reaction times categorizing the pairing of an advantaged group with a positive category and the pairing of a disadvantaged group with a positive category are compared. Generally, reaction times in classifying the disadvantaged group with a positive category are slower than reaction times in classifying the advantaged group with a positive category. This indicates a preference associating the advantaged group with the positive category.

IATs for gender and science reliably find stronger associations between male and science than women and science (Nosek et al., 2002b). This study examined the relationships among gender group membership, strength of identification with the group, and math attitudes and stereotypes. The Nosek et al. (2002b) study used implicit measures, such as IATs, as well as explicit measures, such as self-reports of SAT (i.e., Scholastic Aptitude Test) scores. The researchers found that even though most participants held negative views about math, “women, compared with men, had stronger negative evaluations of math” (Nosek et al., 2002b, p. 50). Additionally, the researchers found that stronger male gender identity was linked to stronger math identity but was not associated with math attitudes; stronger female gender

identification was linked to more negative math attitudes. Interestingly, both men and women classified math and male more easily than math and female.

Automatic associations are found in stereotyping, stereotype threat, and categorization. A key idea is that “concepts that are automatically associated with the self ought to be liked more than should concepts that are less associated with the self” (Nosek et al., 2002b, p. 56). This finding suggests that stereotypes may have broader influence than simply their effects on math performance. It points to the automatic activation of stereotypes (Devine, 1989). It is implied that individuals can be affected by stereotypes on an unconscious level, particularly when referring to situations or areas that are central to one’s sense of self.

Hypotheses

The current project combined the stereotype threat paradigm (Steele & Aronson, 1995) and work on implicit associations (Nosek et al., 2002b). Frantz et al. (2004) were the first researchers to suggest combining the theory of stereotype threat and implicit association tests. Stereotype threat and implicit attitudes can both exist without awareness. Although much work has been done on stereotype threat in women’s math performance, stereotype threat has not been examined as a possible mechanism for the leaky pipeline of women in science. The present study sought to extend the literature by including faculty to examine links between stereotype threat, social identity, and the leaky pipeline of women in science at different points along the pipeline; most studies have solely made use of college students. Because the leaky

pipeline is a phenomenon that spans beyond the college years, it was important to use a broad sample.

The current project sought to test the general hypothesis that a stereotype threat manipulation will result in changes in performance on an implicit association test. The proposed project tested the following hypotheses:

1. In relation to gender, an experimental manipulation of stereotype threat will produce an interaction between experimental condition and gender such that women in the stereotype threat condition will have the highest reaction times on an Implicit Association Test between gender and science.

2. In relation to discipline, an experimental manipulation of stereotype threat will produce an interaction between experimental condition and discipline such that individuals in the sciences who are in the stereotype threat condition will have the highest reaction times on an Implicit Association Test between gender and science.

3. In relation to students and faculty, an experimental manipulation of stereotype threat will produce an interaction between experimental condition and student-faculty status such that faculty in the stereotype threat condition will have the highest reaction times on an Implicit Association Test between gender and science.

4. A prediction model will reveal that variables of gender, discipline, student-faculty status, experimental condition, interest in science, science ability, confidence in science ability, identification with science, overall ability, confidence in overall ability, satisfaction with major/discipline,

persistence in major/discipline, science persistence, and science resilience will be significant predictors of performance on the Implicit Association Test.

The four hypotheses address various aspects of stereotype threat as the theory relates to women in science. The first hypothesis addresses the group membership aspect of stereotype threat (Steele, 1997; Steele & Aronson, 1995) by looking at the effect of gender. This hypothesis is based on the idea that members of a group (i.e., women) are susceptible to stereotype threat because a negative stereotype exists about that group (Steele, 1997; Steele & Aronson, 1995). Therefore, it is expected that women in the stereotype threat condition will show greater endorsement of traditional gender-science stereotypes.

The second hypothesis addresses identification with the stereotyped domain (Steele, 1997) by looking at the effect of identification with science. This hypothesis examines domain identification; individuals who are strongly identified with a domain (i.e., individuals in science disciplines) are susceptible to stereotype threat (Steele, 1997). Therefore, it is expected that individuals in science disciplines in the stereotype threat condition will show greater endorsement of traditional gender-science stereotypes.

The third hypothesis also addressed identification with the stereotyped domain; additionally, the leaky pipeline may be addressed by looking at the effect of student-faculty status. This hypothesis was similar to the second hypothesis in that it examined identification with a domain; individuals who are strongly identified with a domain (i.e., faculty) are susceptible to stereotype threat (Steele, 1997). This hypothesis also addresses the leaky pipeline by examining individuals at various levels of the pipeline.

Therefore, it is expected that faculty in the stereotype threat condition will show greater endorsement of traditional gender-science stereotypes.

The fourth hypothesis was a possible integration of group membership, domain identification, and an implicit association outcome. The theoretical background drew on Schmader, Johns, and Forbes' (2008) conceptual model of the relationships between the group, ability domain, and the self and the individual differences in group identification, domain identification, and stereotype endorsement. The prediction model also drew on Steele's (1997) assertion that stereotype threat can lead to disidentification with the stereotyped domain. Figure 2 displays the proposed model.

CHAPTER 4

METHOD

Participants

Recruitment took place at a large university. A sample was drawn from all students and faculty. Recruitment occurred through announcements at the undergraduate and graduate student senates, and emails to listservs. Faculty members were targeted through fliers individually sent through campus mail, an announcement to the Faculty Senate, and requests to department chairs. A broad range of participants was targeted, including undergraduate and graduate students and faculty at all levels. All participants were 18 years of age or older. Participants included both men and women. Racial/ethnic diversity was considered but is not a primary variable of interest. At the discretion of their instructors, undergraduate student participants had the option of receiving extra credit for participation in research.

In total, 311 participants completed the study. Examining gender, there were 201 (64.6%) female participants and 110 (35.4%) male participants. With respect to race, the sample was fairly homogenous. There were 275 (88.5%) White participants, 12 (3.9%) Asian participants, 6 (1.9%) Black participants, 9 (2.9%) Hispanic participants, and 9 (2.9%) participants who were more than one race, a race other than the races listed above, or did not indicate their race. As for discipline, there were 86 (27.7%) individuals in STEM fields, 85 (27.3%) individuals in social and behavioral science (SBS) fields, 131 (42.1%) individuals in non-STEM fields, and 9 (3%)

individuals who did not indicate their field. As for occupation, there were 238 (76.5%) undergraduate students, 29 (9.3%) graduate students, 16 (5.1%) full professors, 14 (4.5%) associate professors, 9 (2.9%) assistant professors, and 5 (1.6%) adjunct or other faculty. The mean age for all participants was 25.33 years.

Since experimental condition (i.e., stereotype threat activated versus not activated) was randomly assigned, there should have been an approximately equal number of participants in the experimental and control groups. There were 155 participants in the experimental group and 156 participants in the control group.

Measures

An Implicit Association Test (Greenwald et al., 1998) was used (see Appendix A for the instructions that appeared for each block and Appendix B for the selected words/categories used in the task). This task presented words one at a time on a computer screen and required the participant to quickly classify the words into one of two categories. Participants were instructed to press an assigned key on the keyboard depending on which category the word belongs to (i.e., press the “E” key if the word is male or press the “I” key if the word is female). A meta-analysis on 50 IAT studies found an average reliability of .79 (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Similarly, the median reliability across eight IATs was .81 (Nosek & Smyth, 2007). Items used in IATs are generally easily categorized. The male and female names were taken from the names used in Greenwald et al.’s (1998) study on the development of the Implicit Association Test. The science and humanities words were taken from lists in Nosek and Smyth’s (2007) study on the validation of the IAT.

The Gender-Science IAT measured associations between gender and science. Categorization tasks asked participants to respond quickly in classifying items in seven blocks. Table 2 shows the pairings for seven blocks based on the IAT-block paradigm presented in Lane et al.'s (2007) summary of the IAT. The pairings include: 1) male and female names, 2) science and humanities, 3 and 4) male names OR science disciplines and female names OR humanities disciplines, 5) humanities and science, 6 and 7) male names OR humanities disciplines and female names OR science disciplines. In the combined blocks (Blocks 3-4 and Blocks 6-7), the order of the combined pairings was counterbalanced so that order effects did not interfere with the overall IAT effect. Table 2 also shows the number of trials within each block. A trial is the presentation of one word. For each trial, the response latency (i.e., the time needed to classify the word) was measured in milliseconds. Blocks 1, 2, 3, 5, and 6 had 20 trials; blocks 4 and 7 had 40 trials. The combination of 20-trial blocks and 40-trial blocks has been shown to yield good psychometric properties of the IAT (Greenwald et al., 1998; Nosek, Greenwald, & Banaji, 2005; Lane et al., 2007). Although the 20-trial blocks can be regarded as practice, these blocks produce good data (Greenwald et al., 2003). Additionally, there is a larger effect in the first set of combined pairings (Nosek et al., 2005). For example, if participants complete the stereotype-consistent pairing first, the D score will show greater implicit endorsement of the traditional stereotype (i.e., stronger association of science and male than science and female). This is true of all IATs but is particularly noticeable in Gender-Science IAT (Nosek et al. 2005). Having 40 trials in the second block of each pairing reduces the effect of the order of the pairings.

Several short surveys were included in various parts of the study (see Appendix B). A demographic survey included questions about gender, age, ethnicity, occupation, and major/discipline. Ethnicity was included in order to determine whether the sample contained an appropriate representation of minorities. However, ethnicity was not a main variable of interest. An Identification survey, specifically constructed for this study, included questions about interest in math and science, math/science and overall ability, confidence in math/science and overall abilities, identification with math/science, satisfaction with major/discipline, persistence in major/discipline, persistence in math/science, and resilience in math/science. All questions were rated on a 5-point scale where 1 was the lowest score and 5 was the highest score. A brief Familiarity with IATs survey asked how many previous IATs have been taken and awareness of what the measure is testing. Since the IAT measures unconscious associations, previous experience does not necessarily hold a threat of practice effects. The questions were based on the survey used on Project Implicit's demonstration website (Project Implicit, 2007).

Procedures

Participants were instructed to go to a website (during the active running of the study, the website was <http://research.millisecond.com/stammka/iat4.web>) to complete the study (see <http://implicit.harvard.edu> for a demonstration of such a website, as accessed on March 31, 2007). An experiment was designed using Inquisit Desktop Edition (Inquisit, 2008). This program allows the user to custom design web-based experiments, including instruction pages, Implicit Association Tests, surveys, and

randomization. Previous research (Nosek et al., 2002a) has made successful use of web-based demonstrations. Web licenses were purchased to host the experiment online using Inquisit Web Edition.

Participants were randomly assigned to one of four groups. Four groups were necessary in order to counterbalance two order pairings (stereotype-compatible pairings first and stereotype-incompatible pairings first) within each condition to rule out order of presentation of the pairings. The rationale for counterbalancing the ordering of the pairings is similar to the rationale for including more trials in the second set of pairings. The order of the pairings interferes with the IAT effect (Nosek et al. 2005). Counterbalancing helps to reduce the order effects. Groups 1 and 3 were the experimental groups. In group 1, the stereotype-compatible pairings (pair science with male and humanities with female) were presented first. In group 3, the stereotype-incompatible pairings (pair science with female and pair humanities with male) were presented first. Groups 2 and 4 were the control groups. In group 2, the stereotype-compatible pairings (pair science with male and humanities with female) were presented first. In group 4, the stereotype-incompatible pairings (pair science with female and pair humanities with male) were presented first. The ordering of the pairings was counterbalanced. In data analysis, groups 1 and 3 were collapsed into the experimental group, while groups 2 and 4 were collapsed into the control group.

Groups were randomly assigned based on subject ID. Upon beginning the study, a random number from 0 to 1,000,000,000 was generated. Based on the random number, every fourth participant was assigned to a different group. For example, participants with subject ID numbers 1, 5, 9, and so on were in group 1. Participants

with subject ID numbers 2, 6, 10, and so on were in group 2. Participants with subject ID numbers 3, 7, 11, and so on were in group 3. Participants with subject ID numbers 4, 8, 12, and so on were in group 4. This process assured a roughly equal number of participants in each of the groups.

In the experimental condition, participants filled out the demographic survey and then received instructions that the task was intended to measure scientific thinking. This manipulation was designed to increase the salience of gender and gender stereotypes, thereby activating stereotype threat in members of the stereotyped group. Participants then completed the Gender-Science IAT. Following the IAT, participants completed the Identification Survey, and the Familiarity with IATs survey. In the control condition, participants completed the IAT and then completed the Identification Survey, and the Familiarity with IATs survey, and the demographic survey. Participants did not receive instructions about the nature of the task.

Table 3 outlines the procedures and experimental design. The main independent variables were gender (male or female), discipline (science or non-science), student-faculty status (students or faculty), and experimental condition (stereotype threat or no stereotype threat). The main dependent variable was performance on the IAT task. Conventional methods of scoring the IAT involved the computation of a difference score between the response latencies on the two combined tasks (i.e., the mean difference between the response latency on the male+science, female+humanities block and the response latency on the male+humanities, female+science blocks). Greenwald, Nosek, and Banaji (2003) recommend the use of a new algorithm for scoring IAT data that involves the computation of a statistic called

the D score, described next. The D score is “the difference in average response latency between the IAT’s two combined tasks ... divided by an ‘inclusive’ standard deviation of subject response latencies in the two combined tasks” (Lane et al., 2007, pp. 91-92). The D score is loosely analogous to Cohen’s d measure of effect size; the D score provides information about the direction and magnitude of the effect. Calculating the D score involves seven steps (Greenwald, Nosek, & Banaji, 2003; Lane et al, 2007). First, trials with latencies greater than 10,000 milliseconds (ms) were deleted. Second, participants who responded very quickly (more than 10% of trials with latencies less than 300 ms) were deleted. This step is recommended by Greenwald et al. (2003) in order to produce the most psychometrically sound D score. Third, two “inclusive” standard deviations were calculated. The “inclusive” standard deviation is the standard deviation for the trials in blocks 3 and 6 and for the trials in blocks 4 and 7. Fourth, the mean latency was calculated for each of the four blocks. Fifth, mean differences were calculated (mean of block 6 – mean of block 3; mean of block 7 – mean of block 4). Sixth, each mean difference was divided by its own “inclusive” standard deviation. Seventh, the calculations resulted in a D score, which is an adjusted difference score that takes variability into account. The D score represents the male-science association, which refers to how strongly individuals associate male with science. Therefore, a higher score indicates a stronger association with male-science than female-science and greater endorsement of the traditional gender stereotype that science is male.

CHAPTER 5

RESULTS

As a preliminary step in the overall analysis, frequencies, and descriptive statistics, and continuous variables were screened for violations of normality. No special problems with normality were observed. As a preliminary step in preparation of the IAT data, response latencies and error rates on the IAT were examined. Participants with very fast response latencies, as indicated by an average latency of 300 milliseconds or less, or high error rates, as indicated by at least 50% errors, were deleted. In general, participants who had very fast response latencies also made many errors. Three participants were deleted due to very fast reaction times or high error rates.

Preliminary Analyses

Chi-square tests of independence were conducted on gender, discipline, and student-faculty status. This test determines whether there is a relationship between two categorical variables. The test of independence between gender and discipline was significant, $\chi^2(1, N = 302) = 18.69, p < .001$. This result suggests there is a relationship between gender and discipline. There were more men in STEM disciplines as compared to the number of men in non-STEM disciplines, whereas there were relatively equal numbers of women in STEM and non-STEM disciplines. The test of independence between discipline and student-faculty status was significant, $\chi^2(1, N =$

302) = 9.65, $p = .002$. This result suggests there is a relationship between discipline and student-faculty status, showing that there were more students in STEM and more faculty in non-STEM disciplines. The test of independence between gender and student-faculty status was not significant, $\chi^2(1, N = 311) = 2.41, p = .12$. This result suggests there is not enough evidence of a relationship between gender and student-faculty status. Table 4 displays frequency crosstabulations for gender, discipline, and student-faculty status and the results of the chi-square tests.

Analyses

Several analyses were conducted to test the proposed hypotheses. The first three hypotheses were tested by separate two-way analyses of variance (ANOVA), a commonly used method of assessing mean group differences. Two-way ANOVAs contain two categorical independent variables with two or more levels each and one continuous dependent variable. Each of these analyses included gender, discipline, or student-faculty status as the first independent variable; all ANOVAs included experimental condition as the second independent variable. The D statistic was used as the dependent variable.

The fourth hypothesis was tested by a prediction model that was analyzed using structural equation modeling (SEM) techniques. SEM is a set of multivariate methods that can be used to investigate complex phenomena. In general, SEM combines the testing of measurement and structural prediction. SEM can include latent, or unmeasured, factors. Latent factors are indicated by items or subscales from a particular measure. Manifest, or measured, variables can also be used. SEM provides techniques for simultaneously testing multiple hypotheses about latent constructs

(Kline, 2005). In addition to manifest variable and latent constructs, SEM can test mediational relationships. A mediator is a variable that must come between the independent factor and dependent factor (Baron & Kenny, 1986). In order to determine mediational effects, it is necessary to test whether the independent factor is correlated with the mediator, which in turn must correlate with the dependent factor (Collins, Graham, & Flaherty, 1998). A simple correlation between the mediator and the dependent factor is not enough to indicate mediation.

SEMs are evaluated differently than conventional group difference and prediction methods. In SEM, the interest lies in model fit; the chi-square statistic is used to assess model fit. Unlike traditional hypothesis testing, the researcher hopes to find a non-significant chi-square value (i.e., $p > .05$). Ideally, the chi-square value is close to the model's degrees of freedom. A non-significant chi-square value indicates that the data fits the model well, while a significant chi-square value indicates some sort of model misfit. Although the chi-square value gives an indication of overall model fit, the comparative fit index (CFI; Bentler, 1990) provides a more specific assessment of model fit. Ideally, the CFI value should be large (i.e., $> .95$) (Bentler & Hu, 2002). Root mean square error of approximation (RMSEA; Steiger & Lind, 1980) is another measure of fit. Preferably, RMSEA should be small (i.e., ideally $< .05$, or at least $< .10$), indicating little error or discrepancy between a hypothesized model and the sample data. R^2 values were used to assess the amount of shared variance between a mediator or a dependent variable and the variable's direct and indirect predictors. Generally, an R^2 value of .02 is a small effect, an R^2 value of .13 is a medium effect, and an R^2 value of .26 is a large effect.

The model included several categorical independent variables, two mediator constructs, and a dependent variable. Dummy-coded variables were used to represent dichotomous categorical variables (gender: 0 = male, 1 = female; discipline: 0 = non-STEM, 1 = STEM, student-faculty status: 0 = undergraduate students, 1 = graduate students and faculty; experimental condition: 0 = control group, 1 = experimental group). Dummy-coded variables allow the researcher to make interpretations about the relationships between the group membership (i.e., the category coded as 1) and continuous variables (Cohen, Cohen, West, & Aiken, 2003). For example, a positive correlation between gender (where 1 = female) and science identification would indicate that women have higher science identification than men. Conversely, a negative correlation between gender and science identification would indicate that women have lower science identification than men. Two latent constructs of General Academics and Science Identification were used as mediators. Performance on the IAT, as measured by the D score, was the dependent variable.

Gender

A two-way ANOVA was conducted to explore the effects of experimental condition and gender on reaction time on the IAT. This analysis was intended to test the first hypothesis that an experimental manipulation of stereotype threat would produce an interaction between experimental condition and gender. It was anticipated that the overall F ratio, main effects of gender and experimental condition, and an interaction between gender and experimental condition would be significant (i.e., $p < .05$). Effect sizes, as measured by partial eta squared (η^2), were examined. In general, eta squared is the proportion of the total variability in the dependent variable that can

be explained by the independent variable(s). Typically, .01 is considered a small effect, .06 is a medium effect, and .13 is a large effect. It was also expected that women in the stereotype threat group would have the longest reaction times on the IAT. The stereotype threat manipulation was not expected to have a large effect on men because the stereotype does not pertain to men in science domains; in other words, men's performance will not be affected by stereotype threat.

The analysis revealed that there were no significant effects of experimental condition or gender. Additionally, there was no significant interaction between experimental condition and gender. Effect sizes, as measured by partial eta squared (η^2), were all very small (i.e., .004 or lower). Since there were no significant main effects or a significant interaction between experimental condition or gender, post-hoc tests were not conducted. Table 5 shows the D score means and standard deviations for Gender by Experimental Condition, and Figure 3 displays a means plot. Table 6 displays the two-way ANOVA source table for Gender by Experimental Condition.

Discipline

A two-way ANOVA was conducted to explore the effects of experimental condition and discipline on reaction times on the IAT. The second analysis tested the hypothesis that an experimental manipulation of stereotype threat would produce an interaction between experimental condition and discipline. It was anticipated that the overall F ratio, main effects of discipline and experimental condition, and an interaction between discipline and experimental condition would be significant (i.e., $p < .05$). Effect sizes, as measured by partial eta squared (η^2), were examined. It was also expected that individuals in science disciplines in the stereotype threat group

would have the longest reaction times on the IAT. The stereotype threat manipulation was not expected to have a large effect on individuals in non-science disciplines.

Participants were asked to indicate their major or department. From this variable, participants were divided into groups based on whether they were in STEM and non-STEM departments. The determination of STEM and non-STEM departments followed the same breakdown as the NSF ADVANCE Program at the University of Rhode Island (see Appendix D for the full list of STEM and non-STEM departments). For participants who indicated more than one major or discipline, only the first major or discipline was coded. Participants who were undecided were not coded into STEM or non-STEM disciplines. Participants from STEM departments included: Engineering (civil, electrical, computer, and biomedical, and ocean), Environmental and Natural Resource Economics, Biological Sciences, Mathematics, Oceanography, Natural Resource Sciences, Geosciences, Plant Sciences, Computer Science, FAVS, and Cell and Molecular Biology. Participants from non-STEM departments include: Kinesiology, Nursing, Nutrition, Clinical Lab Science, Pharmacy, Education (elementary and secondary), Communication Studies, Business (accounting and finance), Human Development & Family Studies, Political Science, History, Speech Pathology, English, Library, Military Science, Writing, Women's Studies, Theatre, Textile, Fashion Merchandising, & Design, Film, Economics, and German.

The analysis revealed that there were no significant main effects of experimental condition or discipline. Additionally, there was no significant interaction between experimental condition and discipline. Effect sizes, as measured by partial eta squared (η^2), were all very small (i.e., .003 or lower). Since there were no significant

main effects or interaction between experimental condition and discipline, post-hoc tests were not conducted. Table 5 shows the D score means and standard deviations for Discipline by Experimental Condition, and Figure 4 displays a means plot. Table 6 displays the two-way ANOVA source table for Discipline by Experimental Condition.

Student-Faculty Status

A two-way ANOVA was conducted to explore the effects of experimental condition and student-faculty status on reaction times on the IAT. The third analysis tested the hypothesis that an experimental manipulation of stereotype threat would produce an interaction between experimental condition and student-faculty status. It was anticipated that the overall F ratio, main effects of student-faculty status and experimental condition, and an interaction between student-faculty status and experimental condition would be significant (i.e., $p < .05$). Effect sizes, as measured by partial eta squared (η^2), were examined. It was also expected that faculty in the stereotype threat group would have the longest reaction times on the IAT, possibly due to increased exposure to gender stereotypes or increased identification with science domains.

Participants were divided into two groups: students (undergraduate and graduate) and faculty (assistant professors, associate professors, full professors, adjunct faculty, and other faculty). The analysis revealed that there were no significant effects of experimental condition or student-faculty status. Additionally, there was no significant interaction between experimental condition and student-faculty status. Effect sizes, as measured by partial eta squared (η^2), were all very small (i.e., .005 or

lower). Since there were no significant main effects or interaction, post-hoc tests were not conducted. Table 5 shows the D score means and standard deviations for Student-Faculty Status by Experimental Condition, and Figure 5 displays a means plot. Table 6 displays the two-way ANOVA source table for Student-Faculty Status by Experimental Condition.

Revised Discipline and Student-Faculty Status Analyses

Because the planned analyses for discipline and student-faculty status did not reach significance and the original grouping was not considered definitive, revisions to the grouping variables were considered. The revision to the discipline variable was to analyze social and behavioral sciences (SBS) disciplines separately from STEM disciplines. For example, the NSF ADVANCE benchmark indicators are analyzed by STEM and SBS disciplines (ADVANCE Institutional Transformation Indicators Working Group, 2005). Due to the large number of psychology majors in this sample, SBS departments were analyzed as a separate group. In addition to the rationale provided by ADVANCE, there are other reasons to analyze SBS disciplines separately from STEM disciplines. First, SBS disciplines tend to have more women than STEM disciplines. Second, there is a perception that SBS disciplines are not part of the sciences. Third, within SBS disciplines, there is a heightened awareness about gender equity issues. Participants from SBS disciplines include: 1) psychology and 2) sociology & anthropology.

The revision to the student-faculty status variable was to create a new variable with two levels: undergraduate students and all others (graduate students, assistant

professors, associate professors, full professors, adjunct faculty, and other faculty). This revised student-faculty status variable more accurately grouped the more involved individuals (i.e., faculty and graduate students) together, allowing a more fair comparison with less involved undergraduate students. The student-faculty revision was also partly statistical due to very unequal group sizes. The faculty group size was small (i.e., $N = 44$) in comparison to the student group (i.e., $N = 267$). The revised graduate student and faculty group was still small (i.e., $N = 73$) in comparison to the undergraduate student group (i.e., $N = 238$). By moving graduate students into the faculty group and creating slightly more equal group sizes, statistical power was improved. The revision also took into account that some graduate students will decide to pursue academic careers during their graduate studies (van Anders, 2004); some graduate students enter graduate school specifically because they are interested in obtaining jobs in academia (Morgeson et al. 1999).

As a preliminary step, chi-square tests of independence were conducted on gender, revised discipline, and revised student-faculty status. The test of independence between gender and revised discipline was significant, $\chi^2(2, N = 302) = 68.19, p < .001$. This result suggests there is a relationship between gender and revised discipline. There were more women in non-STEM and SBS disciplines and more men in STEM disciplines. The test of independence between gender and revised student-faculty status was marginally not significant, $\chi^2(1, N = 311) = 3.642, p = .056$. This result suggests there is no relationship between gender and revised student-faculty status. The test of independence between revised student-faculty status and revised discipline was not significant, $\chi^2(2, N = 302) = .10, p = .95$. This result suggests that there is no

evidence of a relationship between revised student-faculty status and revised discipline. Table 7 displays frequency crosstabs for gender, revised discipline, and revised student-faculty status and the results of the chi-square tests.

A two-way ANOVA was conducted to explore the effects of revised discipline and experimental condition. The revised discipline variable included three groups: STEM, non-STEM, and SBS. There was a significant interaction between discipline and experimental condition, $F(2, 296) = 4.173$, $p = .016$, partial $\eta^2 = .027$ (small effect). Post-hoc simple effects tests were conducted in order to further understand the interaction. Simple effects tests examine the effects of each group of one independent variable separately over the levels of the other independent variable. The simple effects tests looked at the effects of experimental group over the levels of SBS. For the control group, there were significant differences, $F(2, 149) = 4.621$, $p = .011$, partial $\eta^2 = .058$ (medium effect). Follow-up Tukey tests revealed that STEM ($M = .41$) had a higher mean D score than SBS ($M = .12$). There were no significant differences between 1) non-STEM and STEM and 2) non-STEM and SBS. For the experimental group, there were no significant differences across the levels of SBS. There were no significant effects of experimental condition or discipline. Table 8 shows the D score means and standard deviations for revised Discipline by Experimental Condition, and Figure 6 shows the means plots for the revised Discipline analysis. Table 9 displays the two-way ANOVA source table for the revised Discipline and Experimental Condition analysis.

A two-way ANOVA was conducted to explore the effects of revised student-faculty status and experimental condition. There was a main effect of student-faculty

status, $F(1, 307) = 5.439$, $p = .02$, partial $\eta^2 = .017$ (small effect). Graduate students and faculty ($M = .42$) had a higher D score than undergraduate students ($M = .18$). There was no significant effect of experimental condition. Additionally, there was no significant interaction between experimental condition and student-faculty status. Table 8 shows the D score means and standard deviations for revised Student-Faculty Status by Experimental Condition, and Figure 7 shows the means plots for the revised student-faculty analysis. Table 9 displays the two-way ANOVA source table for the revised Student-Faculty by Experimental Condition analysis.

Prediction Models

The ten items from the Identification Survey were used as independent or mediator variables in the prediction model. The D score was used as the dependent variable. As preliminary steps, correlations were computed between all items. The correlations revealed some high correlations (i.e., .70-.90) and raised some concerns about multicollinearity. Multicollinearity occurs when two variables are highly correlated and may be measuring the same concept (Tabachnick & Fidell, 2007). Although some of the correlations were high, none of the correlations were above .90. However, there was still the possibility of potential collinearity and a reduced chance of predictors reaching significance. Table 10 displays a correlation matrix for the items from the Identification Survey and the D Score for the full sample.

Because the D score had correlations close to zero for almost all of the items, correlations were also computed separately for men and women. Overall, the correlation matrices for men and women were similar, with the exception of the

correlations between the D score and the science items. For men, these correlations were positive, suggesting that as the D score increases, science identification also increases. For women, these correlations were negative, suggesting that as the D score increases, science identification decreases. In the full sample, this pattern of correlations is not discernable. The different pattern of correlations for men and women is consistent with the theory of stereotype threat. As men are seen as being more connected to science, it is understandable that men who believe that science and male are related would have a stronger identification with science, as men. Conversely, women who see a strong connection between being male and science are *less* apt to identify with science, as women. Table 11 displays the correlation matrix for the items in the Identification Survey and the D score for males, and Table 12 displays the correlation matrix for the items in the Identification Survey and the D score for women.

An exploratory factor analysis was conducted in order to preliminarily examine the measurement portion to be used in the structural model. The factor analysis used the 10 items from the Identification survey on the full sample. In determining the number of factors, the number of eigenvalues greater than 1 and the percentage of variance were examined. The factor analysis revealed that there were two factors. All items had high loadings on their respective factors. The first factor, Science Identification, had an eigenvalue of 5.41 and accounted for approximately 54% of the variance. The Science Identification factor included six items: interest in science, science ability, confidence in science ability, identification with science, persistence in science, and resilience in science. The second factor, General

Academics, had an eigenvalue of 1.70 and accounted for approximately 17% of the variance. The General Academics factor included four items: overall academic ability, confidence in overall academic ability, satisfaction with major/discipline, and persistence in major/discipline. Together, the two factors explained approximately 71% of the variance. Factor analyses were also conducted separately for men and women. The results were identical to the analysis using the full sample and indicated the same measurement structure for both men and women.

Reliability for the 10 items was very high (Cronbach's $\alpha = .90$), revealing that participants responded in a consistent fashion to the set of items. Reliability was also computed for the two subsets of Science Identification items and General Academics items, separately. Reliability for the Science Identification items was very high (Cronbach's $\alpha = .94$); reliability for the General Academics items was reasonable (Cronbach's $\alpha = .78$). Table 13 displays the means, standard deviations, and reliability estimates for the Identification Survey items.

A hierarchical multiple regression (MR) was used to preliminarily examine the prediction portion of the structural model. Hierarchical MR allows the researcher to theoretically order the variables into steps. By comparing the change in variance, the researcher can determine how much variance in the dependent variable is accounted for by each additional step. The hierarchical MR was a preliminary step in building a prediction model. The first step included several demographic variables. In the first step, the control variables of gender, discipline, student-faculty status, and experimental condition, were entered. The second step included the four items from the General Academics factor from the Identification survey. In the second step,

overall academic ability, confidence in overall ability, satisfaction with major/discipline, and persistence in major/discipline were entered. The third step included the six items from the Science Identification factor from the Identification survey. In the third step, science interest, science ability, confidence in science ability, science identification, science persistence, and science resilience were entered. The dependent variable was the D score.

None of the steps were significant, although the second step was marginally not significant ($p = .055$). Because of the possible canceling out effect due to positive correlations for men and negative correlations for women, the gender variable was removed from the model, and the regression model was analyzed separately by gender. For men, the first step was marginally not significant $F(3, 103) = 2.51$, $p = .063$, and $R^2 = .07$, revealing a small to medium effect size that may have missed significance due to the low power from a relatively small sample of men (i.e., N for men = 107). The second step was significant, $F(7, 99) = 2.42$, $p = .025$, and $R^2 = .15$, indicating a medium effect size. Confidence in overall ability was a significant predictor of the D score. Student-faculty status (i.e., being a graduate student or faculty member) and overall academic ability were marginally not significant. The third step was significant, $F(13, 93) = 3.79$, $p < .0001$, and $R^2 = .35$, demonstrating a large effect size. Confidence in overall ability and interest in science were significant predictors of the D score. Science persistence and science resilience were marginally not significant predictors of the D score. All other variables were not significant predictors of the D score. The change in R^2 indicated that the variables entered in the second step accounted for 8% of the variance in the D score, above and beyond the 7%

of variance explained by the control variables entered in the first step. The change in R^2 indicated that the variables entered in the third step accounted for 20% of the variance in the D score, above and beyond the 15% of variance explained by the variables in the model in the second step. Table 14 displays the hierarchical MR results for the final step, including significant predictors.

For women, the first step was not significant $F(3, 189) = 1.55, p = .203$, and $R^2 = .02$, showing a small effect size. The second step was also not significant, $F(7, 185) = 1.41, p = .205$, and $R^2 = .05$, again revealing a small effect size. The third step was significant, $F(13, 179) = 2.51, p = .004$, and $R^2 = .15$, indicating a medium effect size. Interest in science and science persistence were significant predictors of the D score. Confidence in science ability was a marginally not significant predictor of the D score. All other variables were not significant predictors of the D score. The change in R^2 indicated that the variables entered in the second step accounted for 3% of the variance in the D score, above and beyond the 2% of variance explained by the control variables entered in the first step. The change in R^2 indicated that the variables entered in the third step accounted for 10% of the variance in the D score, above and beyond the 5% of variance explained by the variables in the model in the second step. Table 15 displays the hierarchical MR results for the final step, including significant predictors.

The main focus of the prediction model was a structural equation model. The model's structure was based on the results of the factor analysis and the preliminary hierarchical MR results. This model potentially offers a more comprehensive and integrated model than the model tested with hierarchical MR. The mediational model

was tested using the EQS program (Bentler & Wu, 2002). Four categorical independent variables were included: gender, discipline, student-faculty status, and experimental condition. There were two mediator latent constructs. The first mediator was the General Academics factor. This factor had four measured variables: overall academic ability, confidence in overall ability, satisfaction with major/discipline, and persistence in major/discipline. The factor loading for overall academic ability was fixed at 1 and was not estimated in order to identify the model. The second mediator was the Science Identification factor. This factor had six measured variables: science interest, science ability, confidence in science ability, identification with science, science persistence, and science resilience. The factor loading for science interest was fixed at 1 and was not estimated in order to identify the model. The dependent variable was the D score. Regression paths were estimated from each of the independent variables to each of the mediator factors. Regression paths were also estimated from the General Academic construct to the Science Identification construct and from the Science Identification construct to the dependent variable of the D score.

Results indicated that the model chi-square was significant, $\chi^2(85, N=300) = 357.82$, $p < .001$, CFI = .88, and RMSEA = 0.10, suggesting some misfit between the proposed model and the data. Although the chi-square test was significant, the chi-square test is sensitive to sample size (Kline, 2005). Chi-square tests on large samples, such as the sample used in this study, tend to be significant simply because of the large sample size. However, the other measures of fit also suggested that the model did not fit the data well, as indicated by a CFI value below .90 and an RMSEA value at .10. The measurement structure indicated that the measured variables loaded significantly

on their respective factors. The path from student-faculty status to the General Academic construct was significant. The paths from gender, discipline, and student-faculty status to the Science Identification construct were significant. The path from the General Academic construct to the Science Identification construct was also significant. However, some of the regression paths were not significant. The paths from gender, discipline, and experimental condition to the General Academic construct were not significant. The path from experimental condition to the Science construct was not significant. Finally, the path from the Science Identification construct to the D score was not significant. R^2 values indicated approximately 16% of the variance in that the General Academics mediator was shared with the direct and indirect relationships with the categorical independent variables. Approximately 38% of the variance in the Science Identification mediator was shared with the direct and indirect predictors. Virtually none of the variance in the D score was shared with its direct and indirect predictors. This represents a medium effect of the General Academics mediator, large effect of the Science Identification mediator, and no effect of the D score. Figure 8 displays the mediational model with standardized parameter estimates and R^2 values. The labeled rectangles represent measured variables. These include the control variables of gender, student-faculty status, discipline, and experimental condition. The remaining labeled rectangles are measured variables from the Identification Survey. The ovals represent latent constructs. Each of the latent variables is connected to several measured variables by lines with one-way arrows, which represent factor loadings. The remaining lines with one-way arrows are regression paths. The numbered rectangles represent standardized parameter estimates,

which are either factor loadings or regression coefficients. For example, the rectangular box on the line from the General Academics latent construct to the Overall Academic Ability measured variable is the value of the factor loading. The rectangular box on the line from the General Academics construct to the Science Identification construct is the value of the regression coefficient.

Because the model fit was not up to expected standards (i.e., ideally a CFI of at least .95 and a RMSEA less than .10), revisions to the model were considered. The Wald test was used to identify possible paths that could be dropped in order to improve model fit. The Wald test suggested dropping the following paths: gender, discipline, and experimental condition to the General Academics factor, student-faculty status and experimental condition to the Science Identification factor, and the Science Identification factor to the D score. These were the non-significant paths in the model.

With the exception of dropping the path from the Science Identification factor to the D score, these revisions suggested by the Wald test were made. Even though there were no paths from experimental condition to other variables, experimental condition was retained in order to demonstrate relationships among hypothesized variables while taking the experimental manipulation into account. Because the D score was hypothesized to be an important variable, it did not make sense to drop this path. Similar to the regression analyses, it was possible that the relationship between Science Identification and the D score was positive for men and negative for women. Essentially, combining men and women in the same sample reduced the overall

relationship to close to zero. Therefore, the gender variable was removed from the model, and the revised model was analyzed separately by gender.

For men, results indicated that the mediational model chi-square was significant, $\chi^2(77, N = 107) = 148.68, p < .001$, CFI = .91, and RMSEA = 0.09. Although the chi-square value was significant, which is common with relatively small samples (e.g., $N = 107$) the data fit the model reasonably well, as indicated by the relatively high CFI value and an RMSEA value less than .10. All measured variables loaded significantly on their respective factors. All regression paths were significant. R^2 values indicated that approximately 11% of the variance in the General Academics mediator was shared with its direct and indirect relationships with the categorical independent variables. Approximately 51% of the variance in the Science Identification mediator was shared with its direct and indirect predictors. Approximately 14% of the variance in the D score was shared variance with its direct and indirect predictors. This represents an almost medium effect of the General Academics mediator, large effect of the Science Identification mediator, and medium effect of the D score. Figure 9 displays the mediational model with standardized solutions and R^2 values.

Two comparison models were also tested. The first comparison model was a full model that contained all of the paths from the mediational model and an additional regression path from the General Academics construct to the dependent variable. The full model chi-square was significant, $\chi^2(76, N = 107) = 148.25, p < .001$, CFI = .91, and RMSEA = 0.09. Similar to the mediational model, the model fit the data reasonably well, as indicated by the high CFI and low RMSEA values; although this

full model was not significantly better than the more parsimonious mediational model. All measured variables loaded significantly on their respective factors. All regression paths, with the exception of the path from the General Academics construct to the dependent variable, were significant. R^2 values indicated that approximately 10% of the variance in the General Academics mediator was shared with its direct and indirect relationships with the categorical independent variables. Approximately 51% of the variance in the Science Identification mediator was shared with its direct and indirect predictors. Approximately 15% of the variance in the D score was shared with its direct and indirect predictors. This represents an almost medium effect of the General Academics mediator, large effect of the Science Identification mediator, and medium effect of the D score. Figure 10 displays the full model with standardized solutions and R^2 values.

A second comparison model tested whether paths to and from the mediator constructs were necessary and only included direct effects between the independent variables and outcome. This direct effects model contained an additional regression path from the General Academics construct to the dependent variable but did not contain the mediational paths from the General Academics construct to the Science Identification construct or the mediational path from the Science Identification construct to the dependent variable. The direct effects model chi-square was significant, $\chi^2(78, N = 107) = 201.72, p < .001$, CFI = .85, and RMSEA = 0.12. The data did not fit the model well, as indicated by the large chi-square, relatively low CFI, and somewhat high RMSEA values. All measured variables loaded significantly on their respective factors. Although the regression paths from student-faculty status to

the General Academics construct and from discipline to the Science Identification construct were significant, the path from the General Academics construct to the dependent variable was not significant. R^2 values indicated that approximately 13% of the variance in the General Academics mediator was shared with its direct and indirect relationships with the categorical independent variables. Approximately 19% of the variance in the Science Identification mediator was shared variance with its direct and indirect predictors. Approximately 4% of the variance in the D score was shared variance with its direct and indirect predictors. This represents an almost medium effect of the General Academics mediator, medium-to-large effect of the Science Identification mediator, and small effect of the D score. Figure 11 displays the direct effects model with standardized solutions and R^2 values.

Chi-square difference tests were computed to compare the three models. For the comparison between the mediational model and full model, the chi-square difference test produced $\chi^2(1, N = 107) = .43, p = .49$. For the comparison between the mediational model and direct effects model, the chi-square difference test produced $\chi^2(1, N = 107) = 53.04, p < .001$. For the comparison between the full model and direct effects model, the chi-square difference test produced $\chi^2(2, N = 107) = 53.47, p < .001$. The chi-square difference tests showed that both the mediational model and the full models represented improvements in model fit over the direct effects model. The chi-square difference tests also showed that there was not enough evidence to say there was a significant difference in the mediational and full models. In comparison to the mediational model, the full model had almost identical chi-square and R^2 values and identical CFI and RSMEA values. This indicates similar fit.

However, the path from the General Academics factor to the D score was not significant. This suggests that the full model may not be as good a theoretical explanation of the data as the mediational model. In comparison to the mediational model, the direct effects model had a poorer fit. This is indicated by the larger chi-square, lower CFI, and higher RSMEA values. The mediational model (Figure 9) was selected as the best model to describe the data for men.

Although the mediational model fit reasonably well for men, further post-hoc adjustments to the model were explored based on the results of Lagrange Multiplier tests. The mediational model in Figure 9 was modified with several additional parameters. First, the bold line indicating a loading from the General Academics latent construct to the science ability measured variable was added. The addition of this path resulted in a complex loading for science ability. That is, science ability loaded onto two latent constructs. This path was added because science ability may be part of overall academic ability as well as science identification. Second, dashed lines indicating correlated errors were added between the errors for the measured variables of 1) overall academic ability and confidence in overall ability, 2) satisfaction with major and persistence in major, and 3) persistence in major and science persistence. The correlated errors were added because the pairs of measures tapped similar content.

The post-hoc mediational model chi-square was significant, $\chi^2(73, N = 107) = 121.69$, $p < .001$, CFI = .94, and RMSEA = 0.08. The post-hoc mediational model represented an improvement in model fit over the original mediational model as indicated by a lower chi-square value, higher CFI value, and lower RMSEA value. A chi-square difference test produced $\chi^2(4, N = 107) = 26.99$, $p < .001$, which also

indicated an improvement over the original mediational model. Figure 12 displays the mediational model with post-hoc adjustments for men.

For women, the mediational model was the same as depicted in Figure 9. Results indicated that the mediational model chi-square was significant, $\chi^2(77, N = 193) = 304.60, p < .001$, CFI = .83, and RMSEA = 0.12. The chi-square value was significant, although this is partially due to the sensitivity of the chi-square statistic to large sample sizes. The model did not fit the data well, as indicated by the relatively low CFI value and somewhat high RMSEA value. However, all measured variables loaded significantly on their respective factors. All regression paths, with the exception of the path from discipline to the Science Identification factor, were significant. R^2 values indicated that approximately 18% of variance in the General Academics mediator was shared with its direct and indirect relationships with the categorical independent variables. Approximately 24% of the variance in the Science Identification mediator was shared with and its direct and indirect predictors. Approximately 3% of the variance in the D score was shared with its direct and indirect predictors. This represents a medium-to-large effect of the General Academics mediator, almost large effect of the Science Identification mediator, and small effect of the D score. Figure 13 displays the mediational model with standardized solutions and R^2 values.

Full and direct effects models were also tested with the subsample of women but did not reveal good fit between the data and the model. The full model chi-square was significant, $\chi^2(76, N = 193) = 297.95, p < .001$, CFI = .84, and RMSEA = 0.12. All measured variables loaded significantly on their respective factors. All regression

paths, with the exception of the path from discipline to the Science Identification factor, were significant. The direct effects model chi-square was also significant, $\chi^2(78, N = 193) = 344.95, p < .001$, CFI = .81, and RMSEA = 0.13. All measured variables loaded significantly on their respective factors. Although the regression path from student-faculty status to the General Academics construct was significant, the paths from discipline to the Science Identification factor and from the General Academics factor to the D score were not significant.

Chi-square difference tests were computed to compare the three models. For the comparison between the mediational model and full model, the chi-square difference test produced $\chi^2(1, N = 193) = 6.65, p < .01$. For the comparison between the mediational model and direct effects model, the chi-square difference test produced $\chi^2(1, N = 193) = 47, p < .001$. For the comparison between the full model and direct effects model, the chi-square difference test produced $\chi^2(2, N = 193) = 40.35, p < .001$. The chi-square difference tests showed that the mediational model and the full models were both better at describing the data than the direct effects model. The chi-square difference tests also showed that the full model was better at describing the data than the mediational model, even though the full model was more complex. Due to the relatively poor fit in all of the three models, none of the models were considered as adequately describing the data for women. It is possible that other variables or other factors are needed, or perhaps a different female sample that has a wider range of identification with science.

Because none of the models fit the data particularly well for women, post-hoc adjustments were explored. Out of the three models, the mediational model appeared

to be the best model to describe the data. The mediational model in Figure 13 was modified with the same additional parameters as the post-hoc mediational model for men. First, the bold line indicating a loading from the General Academics latent construct to science ability was added. Similar to the post-hoc adjustments for men, this path was added because science ability may be part of overall academic ability as well as science identification. Second, dashed lines indicating correlated errors were added between the errors for the measured variables of 1) overall academic ability and confidence in overall ability, 2) satisfaction with major and persistence in major, and 3) persistence in major and science persistence. The correlated errors were added because the pairs of measures tapped similar content.

The post-hoc mediational model chi-square was significant, $\chi^2(74, N = 193) = 250.57, p < .001$, CFI = .87, and RMSEA = 0.11. The post-hoc mediational model represented an improvement in model fit over the original mediational model as indicated by a lower chi-square value, higher CFI value, and lower RMSEA value. A chi-square difference test produced $\chi^2(4, N = 193) = 54.03, p < .001$, which also indicates an improvement over the original mediational model. However, even with the post-hoc adjustments, the chi-square, CFI, and RMSEA values did not indicate good model fit. Figure 14 displays the model with post-hoc adjustments for women.

Familiarity with IATs

Several questions asked participants about previous experiences with IATs. There were 17 (5.5%) participants who indicated that they had previously taken an Implicit Association Test. Of those 17 participants, the mean number of previous IATs

completed was 2.27. A series of yes/no questions asked participants what they thought the task measured. There were 97 (31.2%) participants who indicated that they thought the task measured explicit knowledge. There were 239 (76.8%) participants who indicated that they thought the task measured unconscious knowledge. There were 177 (56.9%) participants who indicated that they thought the task measured general knowledge. There were 78 (25.1%) participants who indicated that they thought the task measured something about women scientists. There were 172 (55.3%) participants who indicated that they thought the task measured stereotypes. An open-ended question enabled participants to indicate if they thought the task measured something else. Common responses included reflexes, cognitive abilities, reaction time, categorization of academic subjects, and association of gender with science and humanities.

CHAPTER 6

DISCUSSION

One of the main goals of the study was to examine whether a stereotype threat manipulation influenced implicit associations regarding gender-science stereotypes. Specifically, the stereotype threat manipulation was expected to have a greater effect on women, individuals in STEM disciplines, and faculty. It was anticipated that the D score would be higher in all of these groups in the stereotype threat condition. Higher D scores indicate greater endorsement of the stereotype that science is male; individuals with higher D scores may suffer negative effects from stereotype threat.

Three hypotheses tested whether stereotype threat affects the leaky pipeline of women in science. There were three main variables that correspond to three pieces of stereotype theory: gender, which tested group membership; discipline, which tested identification with the stereotyped domain; and student-faculty status, which tested domain identification by inclusion of individuals at different points in the pipeline. The fourth hypothesis addressed predictors of the D score and the role of science identification.

Hypothesis 1

The first hypothesis predicted that women in the stereotype threat condition would have the highest D scores on the Implicit Association Test. Previous research has demonstrated that women but not men are susceptible to stereotype threat in math

performance (Murphy et al., 2007). The first hypothesis was not supported. There were no main effects of gender, main effects of experimental condition, or interaction between gender and experimental condition on performance on the IAT. In examining effect sizes, there were very small effects (i.e., partial eta squared (η^2) values of .001 for gender, .000 for experimental condition, and .004 for the interaction) for all analyses.

A possible explanation is that asking participants to indicate their gender on the demographic questionnaire before completing the IAT was not sufficient or too subtle to produce stereotype threat effects. Despite previous research that suggested that asking participants to indicate their race was enough to produce stereotype threat effects in African Americans, the same effect was not found for gender. That is, asking participants to indicate their gender was not enough to produce stereotype threat effects in the current study. Including other demographic questions may have led participants to pay less attention to the gender question. A second possibility is that there are, in fact, no true gender differences. In a review of effect sizes across a wide range of outcomes, Hyde (2005) found little evidence of gender differences and concluded that most outcomes were categorized by gender similarities. However, these possibilities should be interpreted cautiously until further work clarifies the nature of the stereotype threat manipulation used in this study.

Additionally, it is possible that the instruction portion of the stereotype threat manipulation was too subtle to activate the stereotype that women do not do well in science. The instructions indicated that the task measured science ability but did not state anything about gender, gender identity, or expected gender differences. A more

explicit manipulation of social identity would increase the chances that an individual's performance would be influenced by stereotype threat (Schmader, 2002). For example, telling participants that the results would be used in a comparison between men and women may be sufficient to create the hypothesized gender differences in performance.

A future study could use different instructions about the nature of the task. For example, some participants in Frantz et al.'s research (2004, Study 1) received instructions that an IAT would measure racial bias, while other participants received instructions that the IAT would measure awareness about cultural attitudes. In further examining gender stereotypes about science, the instructions could emphasize that the test measures math or science performance and usually shows that men perform better than women. Some participants could receive the example instructions above, while other participants could receive instructions that the test measures math performance and usually shows that men and women perform equally well. Another variation of the instructions could include no information about gender differences. Previous research showed that informing women that a test is unbiased and does not show gender differences facilitated women's performance on a calculus test (Good, Aronson, & Harder, 2008). Similar instructions that either framed a math test as showing gender differences or no gender differences were used by Lesko and Corpus (2006). Task instructions produce different effects; performance will differ depending on whether participants are told that the test is diagnostic of gender differences. Highlighting gender differences activates stereotype threat and impairs women's performance, while highlighting gender similarities counteracts stereotype threat and enhances

women's performance. It is possible that the stereotype threat manipulation instructions had the opposite effect of what was intended. Rather than activate stereotype threat, the instructions may have removed stereotype threat by focusing on science instead of gender. The instructions could have been reworded to include both science and gender, such as stating that the task measures science ability and shows that women generally perform worse than men.

Hypothesis 2

The second hypothesis predicted that individuals in STEM disciplines in the stereotype threat condition would have the highest D scores on the Implicit Association Test. The second hypothesis was supported. As originally proposed, the hypothesis was not supported as there were no significant differences between STEM and non-STEM disciplines, no significant differences between the experimental and control groups, and no significant interaction between discipline and experimental group. A revision of the discipline variable removed SBS out of the STEM group into a separate group. The rationale for this regrouping was that, although SBS is sometimes included within STEM, SBS disciplines should be analyzed separately from STEM disciplines. Additionally, the sample included a large number of participants from SBS disciplines.

There was a significant interaction when the revised discipline variable was used. In the no stereotype threat (control) condition, individuals in STEM disciplines have the highest mean D score, followed by individuals in SBS disciplines, and followed by individuals in non-STEM disciplines. That is, individuals in STEM

disciplines have a stronger association of male-science than individuals in non-STEM disciplines. Interestingly, there were no differences among the three groups in the stereotype threat (experimental) condition. Although the results appear to be the opposite of what was predicted, it is important to keep in mind the direction of the D score. In this study, the D score represents the association of male with science (and presumably female with humanities). Higher D scores represent greater endorsement of traditional gender stereotypes (i.e., higher association of male with science). Conversely, lower D scores represent greater endorsement of nontraditional gender stereotype (i.e., higher association of female with science). This pattern is predicted by stereotype threat theory. Steele (1997) contends that individuals who are more strongly identified with a domain are most susceptible to stereotype threat. Presumably, individuals in STEM disciplines (and SBS disciplines to a lesser degree) are more identified with the sciences than individuals in non-STEM disciplines. The results suggest that individuals in STEM disciplines show a greater endorsement of traditional gender-science stereotypes. The finding that there were no significant differences in the stereotype threat condition suggests that the stereotype threat manipulation may have removed stereotype threat instead of activating stereotype threat. A different manipulation could be similar to the instructions to include both science and gender with an emphasis on the diagnostic nature of the task.

Hypothesis 3

The third hypothesis predicted that faculty in the stereotype threat condition would have the highest D scores on the Implicit Association Test. The third hypothesis

was partially supported. As originally proposed, the hypothesis was not supported as there were no significant differences between students and faculty, no significant differences between the experimental and control groups, and no significant interaction between discipline and student-faculty status. A revision of the student-faculty variable moved graduate students out of the student group and into the faculty group. The rationale for regrouping this variable was that graduate students may have developed a sense of belonging to their domain (Herzig, 2006).

There were significant differences when comparing the revised graduate student-faculty group to the undergraduate group. Although there were no main effects of experimental condition and no significant interaction between experimental condition and the revised student-faculty status variable, there was a main effect of student-faculty status. Overall, graduate students and faculty had a higher mean D score as compared to undergraduate students. That is, graduate students and faculty had a stronger association of male-science than undergraduate students. Similar to the domain identification found in discipline, this pattern is predicted by stereotype threat theory. Presumably, graduate students and faculty are more identified with the sciences than undergraduate students. They show greater endorsement of traditional gender stereotypes.

Hypothesis 4

The fourth hypothesis addressed the prediction of the D score, a measure of implicit endorsement of traditional gender stereotypes. Overall, the structural models showed that the model fit the data reasonably well for men. In the subsample of men,

higher science identification was related to higher endorsement of traditional gender-science stereotypes. For women, the model did not fit as well. The relationship between science identification and endorsement of traditional gender-science stereotypes was weaker and in the opposite direction as would be expected in the subsample of men. These findings are consistent with the prevailing gender-science stereotype. Men who identify with science show implicit attitudes consistent with the “science is male” stereotype. On the other hand, women who identify with science show implicit attitudes inconsistent with the “science is male” stereotype. The finding for women suggests that women in science have to reject the prevailing stereotype in order to fit in with their own behavior and identify with science. In looking at the correlation matrices for the full sample and the subsamples of men and women (Tables 10-12), it is important to note that the only negative correlations occur between the D score and the science identification items for women. This pattern of correlations is reasonable, indicating that women who have high identification with science do not necessarily associate science with men only.

The hierarchical MRs were used as a preliminary step in building the structural model. A limitation of the hierarchical MR is that 13 independent variables were included. With that many independent variables, suppression effects may have occurred. In suppression, the true effects are hidden by the addition of other variables (Cohen et al., 2003). That is, the signs of the relationships can change when suppressor variables are included in a MR model. In looking at the hierarchical MR results for men (Table 14), the fact that some of these relationships are negative, whereas the bivariate correlations were largely positive, indicates that there are

suppressor effects probably due to the multicollinearity of including highly similar items. Similarly, in looking at the hierarchical MR results for women (Table 15), suppression effects appear to occur, again most likely due to multicollinearity when including highly similar independent variables.

The structural equation models showed that the measurement structure was significant in all of the models. Items loaded on their respective latent constructs. However, there were some differences in the prediction structure of the models, in that different regression paths were significant for men and women. The model did not fit as well in the subsample of women, compared to the slightly better model fit for the subsample of men. Several reasons can be offered as explanation. First, there are not, in fact, strong differences in the overall fit of the model between subsamples of men and women. In addition to chi-square values, models are judged by CFI and RMSEA values. As previously described, CFI values above .90 and RMSEA values less than or equal to .10 indicate good model fit. These guidelines are rules of thumb for assessing model fit. In comparing model fit for men and women, the CFI and RMSEA are slightly less acceptable in the female sample than in the male sample. Researchers of statistical reform (Kline, 2005; Cumming, 2008) discourage a focus on the exact numbers and significance tests, and rather advocate a focus on the overall pattern of findings. These researchers contend that there is not much difference between values that are just above or just below the rules of thumb. Second, as men are historically more aligned with science, it is reasonable that links between science identification and associating men with science may not be as strong in a sample of women. This is consistent with gender stereotypes. Third, it is possible that there would be a stronger

negative link between science identification and endorsement of traditional gender-science stereotypes in a sample of STEM women. It is worth noting that, in this study, there was a relatively small sample of women from STEM disciplines (i.e., from STEM, $N = 24$; from SBS, $N = 68$). The lack of a relationship between discipline and the Science Identification construct in the mediational model for women (Figure 12) suggests that there may not have been enough women in the sample who identified with science. Future research could test the prediction that there would be a stronger negative relationship between science identification and stereotype endorsement in STEM women. This prediction is consistent with the literature on stereotype threat.

In light of the model misfit, attention was paid to R^2 values. R^2 represents the proportion of variance accounted for by predictor variables, either directly or indirectly. In the mediational model for men, the R^2 values indicated that the General Academics construct, Science Identification construct, and D score shared 11%, 51%, and 14% of the variance, respectively, with direct and indirect predictors. In the mediational model for men, the R^2 values indicated that the General Academics construct, Science Identification construct, and D score shared 18%, 24%, and 3% of the variance, respectively, with direct and indirect predictors. Even though the model for women did not fit very well and the R^2 values were smaller than the R^2 values for men, the R^2 values were reasonable for the General Academics and Science Identification constructs for women. The finding that the D score only accounted for 3% of the shared variance between the D score and its direct and indirect predictors is consistent with the traditional gender-science stereotype that “science is male.”

Comments on the Experimental Manipulation and Other IAT Issues

It is important to comment on the findings that there were no differences between the experimental conditions in any of the analyses. Task difficulty is a necessary condition in order to produce negative effects because of stereotype threat (Quinn & Spencer, 2001; Spencer, Steele, & Quinn, 1999). That is, a task must be sufficiently challenging so that stereotype threat impairs performance on the task. Theories on the mechanisms behind stereotype threat have focused on anxiety and working memory capacity. Stereotype threat increases anxiety (Osbourne, 2006; Osbourne, 2007) and places a strain on working memory (Beilock, Rydell, & McConnell, 2007). Stereotype threat does not impair performance on easy tasks (Spencer et al., 1999). The IAT may not be sufficiently challenging to produce measurable differences between the experimental and control groups.

Another condition that produces negative effects because of stereotype threat is that the task must be related to the stereotyped domain. Most of the previous work on stereotype threat in women has used math performance (i.e., standardized test scores in math, course grades in math, or the number of math problems attempted) as an outcome measure. The Gender-Science IAT may not be specifically related to math performance.

It is interesting to note that the mean D score on the Gender-Science IAT ($M = .24$, $SD = .44$) in the current study was lower than the mean D scores ($M = .42$, $SD = .49$, and $M = .46$, $SD = .49$) found in two samples in Nosek et al. (2005). Both studies used Internet samples. However, there are several key differences between the two studies. First, the sample in the current study was drawn from a university setting,

whereas the samples from Nosek et al. (2005) were drawn from the general population. Second, the data in the current study were collected between March 23, 2009 and May 5, 2009. The data in Nosek et al. (2005) were collected between November 15, 2001 and October 29, 2002. The studies had differences in the timing of data collection. The data in the current study was collected over a shorter time period (i.e., 1.5 months in comparison to 1 year) and in a different era. Even within a few years' difference, there may have been cultural changes in gender-science stereotypes. Third, the sample in the current study was much smaller ($N = 311$), whereas each of the two samples in Nosek et al. (2005) contained about 12,000 participants.

Stereotype Threat and Women in Science

Stereotype threat in women in science can be organized around three themes: ability and choice, identity, and motivation. The first theme, ability and choice, touches on differences in ability and differences in choice. Rosenbloom et al. (2008) cite ability differences and choice, as well as discrimination as possible reasons for under-representation of women in science. A common misconception is that there are gender differences in math ability. However, it is important to note that women's lack of involvement in the sciences is not a result of ability (Ceci et al., 2009; Rosser & Taylor, 2009). A review of cognitive differences found that the genetic basis for math and science reasoning is the same for men and women (Spelke, 2005). Additionally, a recent review of state educational data found that there were no gender differences in math performance (Hyde et al., 2008). Although formerly thought to be an explanation of under-representation of women in science, ability differences are not

likely to be a factor in the leaky pipeline. Emphasizing the malleability of ability (Aronson, Fried, & Good, 2002) or the malleability of implicit stereotypes (Blair, Ma, & Lenton, 2001; Dasgupta & Asgari, 2004) reduces the negative effects of stereotype threat. Emphasizing effort, rather than ability, also reduces the negative effects of stereotype threat (Thoman et al., 2009). Thus, defining ability and implicit stereotypes as malleable or redefining ability as effort may be strategies to overcome stereotype threat. Since differences in ability do not sufficiently explain the under-representation of women in the sciences, choice provides a better explanation for this phenomenon. Choice is related to interest and preference; individuals choose to pursue domains in which they have interests. In this view, interest is related to discrepancies between self-concept and stereotypes about disciplines (Lee, 1998). Individuals pursue domains in which there is a match between their self-concepts and stereotypes. Certain disciplines, such as the sciences, are seen as male. Therefore, women are less likely to pursue interests in male-oriented disciplines because of the mismatch between their self-concepts as women and stereotypes about appropriate domains for women. Ceci et al. (2009) suggest that women's preferences or choices, not biological factors, are the best explanation of the under-representation of women in the sciences (Ceci et al., 2009). In a longitudinal study, Simpkins, Davis-Kean, and Eccles (2008) found that choices and interests in 5th grade were positively related to the number of math and science courses taken in high school. Choices in middle school can lay the foundation for future choices in math and science.

A second key theme in stereotype threat in women in science is the role of identity. The salience of social identity can be influenced by the presence (or absence)

of others (Jordan & Lovett, 2007). In the case of multiple social identities, it also matters which social identity is activated. Shih, Pittinsky, and Ambady (1999) studied the role of stereotype threat in a sample of Asian American women. In regard to math performance, Asian American women face two social identities: ethnicity and gender. These researchers found that activating gender identity impaired women's math performance, while activating ethnic identity enhanced women's math performance. The pattern of impairment and enhancement is consistent with gender stereotypes about math. Sinclair, Hardin, and Lowery (2006) observed a pattern of stereotypic ratings depending on which identity was most salient. When ethnic (i.e. Asian American) identity was activated, women rated their math ability more favorably. This finding is consistent with the stereotype that Asians perform well at math. When gender identity was activated, however, women rated their verbal ability more favorably. This finding is consistent with the stereotype that women do not perform well at math. Research on multiple identities (Gresky et al., 2005) suggests that activating complex social identities, such as mapping the self-concept with many nodes and connections between the nodes, will protect women from the negative effects of stereotype threat. A final strategy is to remind women of their individual identities (Ambady et al., 2004). Like the effects of the availability of multiple identities, individual identity plays a protective role and can prevent the negative effects of stereotype threat.

At least two main consequences of stereotype threat can be noted. The first consequence is distancing from the domain in the form of disidentification (Jordan & Lovett, 2007). That is, women in science who experience stereotype threat are more

likely to move away from science disciplines. This serves to reduce the impact of the threatened identity and to preserve the self-concept. Steele (1992) hypothesized that disidentification with academics occurs in African Americans. Disidentification occurs gradually over an individual's academic career (Steele, 1992). Osbourne (1995) and Cokley (2002) provided some support for the disidentification hypothesis by examining the relationship between self-esteem and academic performance. Results suggested that, over time, African Americans had a pattern of weakening correlations between self-esteem and academic performance. A second consequence of stereotype threat that is related is disengagement from the stereotyped domain (Steele, 1997; Jordan & Lovett, 2007). This is precisely what happens in the leaky pipeline of women in science. Women who experience stereotype threat are likely to not care about science performance. However, there is an ironically adaptive feature of disengagement. In certain circumstances, disengagement results in an increased likelihood of persistence (Nussbaum & Steele, 2007). This process allows for protection of the self in response to stereotype threat. Rather than leave the domain, the individual minimizes the psychological threat and persists in the face of an identity threat.

A third key theme in stereotype threat in women in science is motivation. Motivation can include such concepts as self-efficacy and persistence. Self-efficacy results in "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). According to Bandura's theory (1977), expectations about self-efficacy can be increased by performance accomplishments, that is, a sense of mastery. Individuals become involved in activities

that they are capable of handling and avoid situations that are threatening. This partially explains why individuals who experience stereotype threat may leave a domain. Stereotype threat decreases performance and thereby also decreases self-efficacy expectations. Increasing self-efficacy may result in increased career exploration in male-oriented disciplines (Betz & Schifano, 2000). There is some evidence that persistence can aid in women's pursuit of the sciences. For example, self-efficacy and academic goals were related to persistence (Brown et al., 2008). In light of this finding, persistence may be a good variable to include in a model of science identification. Additionally, coupling beliefs about one's ability to achieve success with future goals and motivations could be an effective strategy for getting more women to stay in the science pipeline.

Implications for the Leaky Pipeline

The results of this study provide some evidence that traditional gender stereotypes influence the leaky pipeline of women in science. Overall, there was no evidence that the experimental manipulation had an effect; and there was no evidence that the gender analyses showed significant differences. However, the discipline and student-faculty analyses provide some evidence that traditional gender stereotypes are present. The discipline analyses showed that, for participants in the control group, individuals in STEM disciplines hold more traditional gender stereotypes than individuals in non-STEM disciplines. Similar to the discipline findings, overall graduate students and faculty hold more traditional gender stereotypes than undergraduate students. Although this study did not link the presence of stereotypes to

actual behavior or future outcomes, the “science is male” stereotype could function in subtle ways to prevent women in science from performing at their best.

The prediction models indicate that science interest, persistence, and resilience are related to science identification. Even though there were some differences between the regression and structural equation models, the prediction models suggest that science ability and resilience may be related to science identification. The models have implications for the leaky pipeline in terms of addressing interest. Although choice contributes to the under-representation of women in science, discrimination also contributes to this phenomenon (Rudman et al., 2008). Discrimination can occur when women violate gender stereotypes, such as when women succeed at tasks that are stereotyped as being male (Heilman & Okimoto, 2007). Additionally, subtle bias and accumulation of disadvantages over time (Valian, 1998) further contribute to discrimination. The surrounding culture can also create bias, such as subtle bias in the media (Davies et al., 2002). Although overt harassment is less acceptable than in the past, sexism continues to exist. Subtle discrimination harms women. For example, women experienced a threat from interacting with sexist men (Logel et al., 2009); and the presence of sexism impaired women’s cognitive performance (Dardenne et al, 2007). In the sciences, women are more likely to encounter sexism and therefore are susceptible to impairment in performance.

In addition to discrimination, pipeline paths are blocked for women, particularly at higher levels of professional advancement (Dominici et al., 2009). Rosser and Taylor (2009) suggest that two ways to unblock these paths are to give more attention to work-life issues and networking (Rosser & Taylor, 2009). Women

face a trade-off between work and family (Ceci et al., 2009), thereby reducing the number of women who remain in the science pipeline. Women who do remain in the science pipeline have fewer networking opportunities in academic leadership positions (Dominici et al., 2009). Marx and Roman (2002) noted the positive effects of female role models who are competent in math. Both networking and role models may remove the barriers to women's advancement in the sciences; and thus are factors that could be important in improving retention.

Creating a culture of inclusive excellence, a setting that empowers students to excel in academics, encourages members of underrepresented groups to pursue science careers (Hrabowski & Maton, 2009). Strategies to reduce prejudice may assist in adding women to science domains. For example, Crisp and Turner (2009) suggest that imagining positive interactions between members of different groups (i.e., men in science and women in science) may help to promote women in the sciences. By doing so, individuals can create positive perceptions of women in science and facilitate successful interactions with women in science.

Valian (2006) stresses the desirable qualities of gender equity, such as maximizing talent in the pool of new faculty, increasing the likelihood of making innovations in teaching, research, and scholarship, and increasing job opportunities for students. It is important to increase the number of women in science, but increasing numbers alone will not ensure that women remain in the sciences. More women need to advance to higher positions in the pipeline in order to serve as role models for upcoming women leaders (Kite et al., 2001). A sufficient number of female models in

leadership positions will help to provide mentoring and support in order to unblock paths to advancement that were previously unavailable to women.

A metaphor to describe the blocked points in the science pipeline is to view the blocks as clogs or constrictions in the pipeline. Depending on which image is selected, the blocks appear different. Clogs can be viewed as immovable barriers that are difficult or impossible to overcome. Hitting a clog creates strong resistance and prevents anything from flowing in the pipeline. On the other hand, constrictions can be viewed as challenges that can be overcome with persistence and continued effort. The women who advance up the science pipeline are the women who press through the narrow points and try again until they succeed. Both the individual and the institution can play a role in women's persistence in the sciences. Persistence occurs on the part of the individual (Herzig, 2006), while retention occurs on the part of mentors or institutions (Herzig, 2006). Women who are more likely to persist in the sciences are more likely to stay in the science pipeline and to advance to higher positions.

Limitations

One limitation concerns the relationship between explicit and implicit attitudes. Research has found that implicit attitudes and explicit attitudes are distinct but related concepts (Greenwald & Farnham, 2000; Nosek, 2005). This study did not measure explicit attitudes toward gender and science. Therefore, it was not possible to compute a correlation between explicit and implicit attitudes. An example of a direct measure of gender and science would be to ask participants to rate their attitudes toward science measured from strongly like to strongly dislike (Nosek et al, 2002a).

Such a question would enable the calculation of a correlation between implicit and explicit attitudes about gender stereotypes.

A second limitation is the small sample of faculty. This is partly due to the smaller number of faculty (approximately 600 tenure-track faculty) in comparison to the number of students (approximately 16,000 undergraduate and graduate students) at the University of Rhode Island (URI Institutional Research, 2008). This is also due to compatibility issues with the software used in this study. The software only ran on computers using Windows; it was not compatible with Macs or other operating systems. Future research could address this limitation by running the study in person or by expanding to faculty beyond the University of Rhode Island.

A third limitation was the measures of identification, ability, persistence, and resilience from the Identification survey. These constructs were measured by single items. Ideally, multiple items should assess each construct. For example, the Occupational Self-Efficacy Scale (Betz & Hackett, 1981) measures self-efficacy for 20 female-oriented (i.e., art teacher) and male-oriented occupations (i.e., engineer). Similarly, Lee (1998) measured rating of interest in becoming a professional in eight science professions, such as biologist and chemist. This type of measure is an alternative way of conceptualizing interest in science. Additional questions on domain identification could include competence, liking of the domain, rewards, and sense of belonging to the domain. These four considerations are based on the domain identification principles outlined in Cullen et al. (2006) and Steele (1997). For students, intent to major in the sciences (Cullen et al., 2006) could be used as a measure of domain identification. The Bem Sex-Role Inventory (Bem, 1974)

measures the degree of masculinity and femininity and could be used to assess the endorsement of gender stereotypes. A future study could include additional questions about achievement orientation, identity and the self-concept.

A fourth limitation was that this project was an exclusively web-based study. This creates two potential problems. First, the sample may not be a representative sample of the population of interest (Martin, 2008). Sampling was limited to participants who owned or had access to computers and who had Internet access. However, this problem occurs in most samples drawn from a college population. Although Internet samples are not representative of the general population, Internet samples do not appear to be different from samples obtained by traditional methods of psychological research (Gosling et al., 2004). Results from web-based samples yield similar results to samples from traditional methods, such as paper-and-pencil surveys (Gosling et al., 2004). Second, like experiments conducted in laboratories, a web-based study is not conducted in a real-world setting. However, other studies (Greenwald et al., 2003; Nosek et al., 2002a; Nosek, 2005, Nosek et al., 2005) have made effective use of web-based research on implicit association.

Finally, a fifth limitation was noted in the Familiarity with IATs survey. Approximately half (55.3%) of the participants responded “yes” when they were asked whether they thought the task measured stereotypes. It is relatively easy to guess the purpose of the IAT (Frantz et al., 2004); it appears that some participants were able to guess the purpose of the IAT in the current study. It is not clear, however, whether participants were able to guess the purpose because of the categories used in the IAT or because of the items in the Identification survey. It is possible that the order of

surveys in the study influenced responses to this question. For all participants, the Familiarity with IATs survey was completed after the IAT and the Identification survey. The items on the Identification survey, particularly the items pertaining to the science, may have allowed some participants to guess the purpose of the study. However, since participants were not asked to specify what they thought the question was measuring, it is not possible to determine how participants interpreted and responded to this question.

Future Directions

Future directions include developing an integrative model of the multiple factors that contribute to stereotype threat and implicit attitudes. For example, structural equation modeling is suited to examine models that combine measurement of latent, or unmeasured, variables with prediction between variables. Nosek and Smyth (2007) proposed a multitrait-multimethod approach to validate the Implicit Association Test. This study found that implicit and explicit attitudes are related but are separate from each other. Additionally, other types of structural equation models allow for the inclusion of mediator and/or moderator variables, such as identification with a stereotyped domain.

Another future direction is to examine what can be done about implicit attitudes. It is often frustrating for individuals to learn that they hold implicit attitudes (i.e., greater preference for male-science associations than for female-science associations), even though the implicit attitudes may not match the individual's explicit attitudes (i.e., publicly believing in gender equality in science achievement).

Attitudes and behavior do not always coincide. That is, an individual may hold an attitude but not act on it. Similarly, a particular behavior does not always indicate that an individual holds a particular attitude. Future research could consider questions such as: under what conditions can implicit attitudes change? Do implicit attitudes lead to discriminatory behavior? What, if any, interventions would be useful to minimize the effects of implicit attitudes? For example, it has been suggested that implicit attitudes include affective processes (Rudman, Ashmore, & Gary, 2001). It has also been suggested that stereotype threat includes affective processes (Lesko & Corpus, 2006). Interventions that target emotions may be beneficial.

A third future direction is to substitute the Gender-Science IAT with a Gender-Career IAT. All other procedures, design, and analyses could replicate the ones found in this proposal. The same lists of male-female names could be used, and a categorization task asking participants to make determinations about “career” or “home” words. A study using a Gender-Career IAT may shed light on additional mechanisms of the leaky pipeline of women in science. Among other factors, a key to women’s successful advancement in science is attention to work-life balance (Rosser & Taylor, 2009 Ceci et al., 2009). Women are more likely to shoulder the burden for child care and household management. Women are also more likely to be negatively impacted by work-life issues (Rudd et al., 2008). One strategy to increase work-life balance is to provide flexible employment patterns, such as part-time jobs and telecommuting (Donovan et al., 2005; Ceci et al., 2009). These types of arrangements may help to alleviate the advancement barriers faced by women.

A fourth future direction is to study stereotype threat in women in science in a real-world setting. Most of the research on stereotype threat on gender stereotypes has been conducted in laboratories. To date, stereotype threat has not been consistently demonstrated in real-world settings. For example, recent studies found that there were no gender differences in the strength of the relationship between SAT-Math scores and grades in college English courses (Cullen, Hardison, & Sackett, 2004; Cullen, Waters, & Sackett, 2006). Stereotype threat theory predicts that only women would be affected by stereotype threat and that the pattern of the relationships for men and women would be different. However, this was not observed in Cullen et al. (2004) and Cullen et al. (2006). This finding is opposite of the prediction of stereotype threat theory. Future work in other real-world settings can demonstrate whether stereotype threat occurs outside of the laboratory.

Conclusions

Although there is still much work to be done to increase the numbers of women in science, it is encouraging that efforts have been taken to repair the leaky pipeline. It is encouraging that interventions that remove stereotype threat can facilitate performance (Good et al., 2008; Lesko & Corpus, 2006). Focusing on science persistence and finding ways to encourage women to use their abilities in science have the potential to keep more women in the science pipeline. Ultimately, persistence, among other factors, translates into retention.

Stereotype threat undermines the self-concept through identity threats based on group membership. Belonging is a theme of social psychology (Fiske, 2004). At its

core, social identity theory explains the need to belong. Stereotype threat undermines the need to belong in groups that are negatively stereotyped, such as women in science. For example, women are less likely to associate themselves with science at an early stage in the science pipeline (Lee, 1998). Stereotype threat also undermines the self-concept through identity threats based on domain identification. Stereotypes dictate who can belong to certain domains. Stereotypes do not present a problem for individuals who fit the stereotype (i.e., men in science), but stereotypes are problematic for individuals who do not conform to the stereotype (i.e., women in science).

Although not likely to occur overtly, stereotype threat occurs in subtle forms. This is precisely why stereotype threat is harmful. Because of subtle bias, indirect measures, such as the Implicit Association Test, can provide a window into whether stereotype threat occurs on an unconscious or automatic level. In general, IATs reveal bad news and good news: we are all prejudiced (Kelley, 2009). In other words, although everyone holds biases, acknowledging that everyone holds biases can lead to solutions to stereotyping. Being aware that a stereotype exists is a step in the right direction. The question becomes what can be done to overcome the effects of stereotype threat.

This research has implications for women who are already in the science pipeline and intend to continue to higher points in the pipeline. Outreach activities, programs to promote women's involvement in the sciences, and awareness of implicit attitudes could be used to ensure advancement of the women who have not yet left the science pipeline. Other activities can shed light on the subtle bias present in stereotype

threat. Future work can address how to prevent the negative effects of stereotype threat and how to minimize or eliminate stereotype threat once it has been activated. The prediction model suggests that gender differences in science identification exist such that men are more strongly identified with science than women. This may create tension for women in science because of a mismatch between the traditional gender stereotype and women's identification with science. Future work can further elaborate the role of science identification in the retention of women in science.

Table 1. Selection of Empirical Literature on Stereotype Threat in Women in Science

Reference	Sample and Type	Area	Findings
Davies et al., 2005	Undergraduate men and women	Identity, social identity, gender stereotypes	Reduced effects of stereotype threat by eliminating stereotype threat vulnerability (i.e., no gender differences on a leadership task)
Kiefer & Sekaquaptewa, 2007	Undergraduate women	Situational cues, implicit associations between gender and math, gender stereotyped, math identification	Stereotype threat and implicit gender-math attitudes interact with each other
Kawakami et al., 2008	Undergraduate women	Identification with math, implicit attitudes toward math	Training to approach math resulted in positive implicit attitudes toward math and higher identification with math
Marx & Stapel, 2006	Undergraduate men and women	Social identity, priming	Stereotype threat creates performance deficits in members of the stereotyped group
Murphy, Steele, & Gross, 2007	Undergraduate men and women	Situational cues, social identity, group membership	Identity threats (i.e., number of women) and situational cues can affect performance in a particular setting
Smith, Sansone, & White, 2007	Undergraduate women	Interest in math/science, achievement motivation, salience of gender	Lower interest, task avoidance, and task engagement are related to stereotype threat

Table 2. Implicit Association Test (IAT) Block Overview

Counterbalancing of Stereotype-Consistent Pairings

		Stereotype-Compatible Pairings First Group		Stereotype-Incompatible Pairings First Group	
Block	# Trials	Left Key Assignment	Right Key Assignment	Left Key Assignment	Right Key Assignment
1	20	male	female	humanities	science
2	20	science	humanities	male	female
3	20	science OR male	humanities OR female	humanities OR male	science OR female
4	40	science OR male	humanities OR female	humanities OR male	science OR female
5	20	humanities	science	science	humanities
6	20	humanities OR male	science OR female	science OR male	humanities OR female
7	40	humanities OR male	science OR female	science OR male	humanities OR female

Note: The first set of blocks presents the stereotype-compatible pairings first (pairing science and male and pairing humanities and female). The second set of blocks presents the stereotype-incompatible pairings first (pairing science and female and pairing humanities and male).

Half of the participants in the experimental and control groups were in each of the counterbalanced conditions.

Table 3. Procedures and Experimental Design

Experimental Condition	# participants	Independent Variables	Dependent Variable	Procedures
Stereotype Threat (Experimental group)	155	<p>One of the following:</p> <ul style="list-style-type: none"> • Gender • Discipline • Student-Faculty Status <p>Plus: Experimental Condition</p>	Performance on IAT	<p>Demographic survey</p> <p>Science instruction</p> <p>Gender-Science IAT task</p> <p>Identification survey</p> <p>Familiarity with IAT survey</p>
No Stereotype Threat (Control Group)	156	<p>One of the following:</p> <ul style="list-style-type: none"> • Gender • Discipline • Student-Faculty Status <p>Plus: Experimental Condition</p>	Performance on IAT	<p>Gender-Science IAT task</p> <p>Identification survey</p> <p>Familiarity with IAT survey</p> <p>Demographic survey</p>

Table 4. Frequency Crosstabulations for Gender, Discipline, and Student-Faculty Status and Chi-Square Results

		Discipline		
		STEM	Non-STEM	Total
Gender	Men	79	29	108
	Women	92	102	194
	Total	171	131	302
		$\chi^2(1, N=302)= 18.69, p<.001$		
		Discipline		
		STEM	Non-STEM	Total
Student-Faculty Status	Students	156	103	259
	Faculty	15	28	43
	Total	171	131	302
		$\chi^2(1, N=302)= 9.65, p=.002$		
		Student-Faculty Status		
		Students	Faculty	Total
Gender	Men	99	11	110
	Women	168	33	201
	Total	267	44	311
		$\chi^2(1, N=311)= 2.41, p= .12$ (not significant)		

Table 5. D Score Means and Standard Deviations for Gender, Student-Faculty Status by Experimental Condition

D score Mean (SD) N = 311		Experimental Condition		
		Stereotype Threat	No Stereotype Threat	Total
Gender	Men	.23 (.45)	.28 (.46)	.25 (.45)
	Women	.26 (.43)	.21 (.44)	.23 (.43)
	Total	.25 (.44)	.24 (.44)	Overall = .24 (.44)
D score Mean (SD) N = 302		Experimental Condition		
		Stereotype Threat	No Stereotype Threat	Total
Discipline	STEM	.24 (.45)	.27 (.45)	.25 (.45)
	Non-STEM	.28 (.43)	.21 (.45)	.24 (.44)
	Total	.25 (.44)	.24 (.45)	Overall = .25 (.44)
D score Mean (SD) N = 311		Experimental Condition		
		Stereotype Threat	No Stereotype Threat	Total
Student-Faculty Status	Students	.25 (.44)	.21 (.44)	.23 (.44)
	Faculty	.24 (.44)	.39 (.39)	.31 (.42)
	Total	.25 (.44)	.24 (.44)	Overall = .24 (.44)

Table 6. Two-way ANOVA Source Table for Gender, Discipline, and Student-Faculty Status by Experimental Condition Analyses

Source	SS	df	MS	F	p	Partial Eta Squared (η^2)
Gender and Experimental Condition						
Gender	.037	1	.037	.193	.661	.001
Experimental Condition	.000	1	.000	.000	.977	.000
Gender x Experimental Condition Interaction	.210	1	.210	1.078	.300	.004
Error	59.664	307	.194			
Total	59.929	310				
Discipline and Experimental Condition						
Discipline	.005	1	.005	.023	.879	.000
Experimental Condition	.029	1	.029	.148	.701	.000
Discipline x Experimental Condition Interaction	.189	1	.189	.966	.327	.003
Error	58.253	298	.195			
Total	58.463	301				
Student-Faculty Status and Experimental Condition						
Student-Faculty Status	.257	1	.257	1.328	.250	.004
Experimental Condition	.101	1	.101	.524	.470	.002
Student-Faculty Status x Experimental Condition Interaction	.316	1	.316	1.633	.202	.005
Error	59.342	307	.193			
Total	59.929	310				
Notes: * = $p < .05$						
Column headings: Source = source of variance, SS = sum of squares, df = degrees of freedom, MS = mean square, F = F-test value, p = probability, partial eta squared (η^2)= effect size/proportion of explained variance						

Table 7. Frequency Crosstabulations for Gender, Revised Discipline, and Revised Student-Faculty Status and Chi-Square Results

		Revised Discipline			
		STEM	Non-STEM	SBS	Total
Gender	Men	62	29	17	108
	Women	24	102	68	194
	Total	86	131	85	302
		$\chi^2(2, N=302)= 68.19, p<.001$			
		Revised Student-Faculty Status			
		Undergraduate Students	Graduate Students & Faculty	Total	
Gender	Men	91	19	110	
	Women	147	54	201	
	Total	238	73	311	
		$\chi^2(1, N=311)= 3.642, p=.056$ (not significant)			
		Revised Discipline			
		STEM	Non-STEM	SBS	Total
Revised Student-Faculty Status	Undergraduate Students	65	100	66	231
	Graduate Students & Faculty	21	31	19	71
	Total	86	131	85	302
		$\chi^2(2, N=302)= .10, p= .95$ (not significant)			

Table 8. D Score Means and Standard Deviations for Revised Discipline and Revised Student-Faculty Status by Experimental Condition

D score Mean (SD) N = 302		Experimental Condition		
		Stereotype Threat	No Stereotype Threat	Total
Revised Discipline	STEM	.19 (.42)	.41 (.45)	.30 (.44)
	SBS	.28 (.47)	.12 (.40)	.20 (.44)
	Non-STEM	.28 (.43)	.21 (.45)	.24 (.44)
	Total	.25 (.44)	.24 (.45)	Overall = .25 (.44)
D score Mean (SD) N = 311		Experimental Condition		
		Stereotype Threat	No Stereotype Threat	Total
Revised Student-Faculty Status	Undergraduate Students	.24 (.44)	.17 (.43)	.21 (.44)
	Graduate Students & Faculty	.27 (.42)	.42 (.43)	.35 (.43)
	Total	.25 (.44)	.24 (.44)	Overall = .24 (.44)

Table 9. Two-way ANOVA Source Table for Revised Discipline and Revised Student-Faculty Status by Experimental Condition Analyses

Source	SS	df	MS	F	p	Partial Eta Squared (η^2)
Revised Discipline and Experimental Condition						
Revised Discipline	.469	2	.234	1.228	.294	.008
Experimental Condition	.001	1	.001	.007	.936	.000
Revised Discipline x Experimental Condition Interaction	.210	2	.210	1.078	.016*	.027
Error	56.494	296	.191			
Total	58.463	301				
Revised Student-Faculty Status and Experimental Condition						
Revised Student-Faculty Status	1.030	1	1.030	5.439	.020*	.017
Experimental Condition	.104	1	.104	.551	.459	.002
Revised Student-Faculty Status x Experimental Condition Interaction	.701	1	.701	3.704	.055	.003
Error	58.132	307	.189			
Total	59.929	310				
Notes: * = $p < .05$						
Column headings: Source = source of variance, SS = sum of squares, df = degrees of freedom, MS = mean square, F = F-test value, p = probability, partial eta squared (η^2)= effect size/proportion of explained variance						

Table 10. Correlation Matrix for Identification Survey Items and D Score for the Full Sample

Item	Science Interest	Science Ability	Confidence in Science Ability	Science Identification	Overall Academic Ability	Confidence in Overall Academic Ability	Satisfaction with Major/Discipline	Persistence in Major	Science Persistence	Science Resilience
Science Ability	.73	--								
Confidence in Science Ability	.70	.87	--							
Science Identification	.78	.80	.77	--						
Overall Academic Ability	.29	.45	.40	.32	--					
Confidence in Overall Academic Ability	.29	.41	.44	.32	.62	--				
Satisfaction with Major/Discipline	.28	.33	.28	.31	.39	.43	--			
Persistence in Major	.25	.32	.26	.23	.46	.40	.59	--		
Science Persistence	.72	.67	.66	.73	.32	.26	.31	.36	--	
Science Resilience	.74	.71	.66	.71	.26	.27	.25	.25	.76	--
D Score	.02	.03	.00	.03	.13	.03	.04	.13	.02	.02

Table 11. Correlation Matrix for Identification Survey Items and D Score for Men

Item	Science Interest	Science Ability	Confidence in Science Ability	Science Identification	Overall Academic Ability	Confidence in Overall Academic Ability	Satisfaction with Major/Discipline	Persistence in Major	Science Persistence	Science Resilience
Science Ability	.77	--								
Confidence in Science Ability	.78	.86	--							
Science Identification	.81	.79	.80	--						
Overall Academic Ability	.41	.56	.43	.39	--					
Confidence in Overall Academic Ability	.32	.49	.46	.35	.61	--				
Satisfaction with Major/Discipline	.38	.41	.38	.41	.45	.41	--			
Persistence in Major	.31	.42	.36	.31	.49	.47	.56	--		
Science Persistence	.56	.60	.59	.67	.45	.29	.44	.51	--	
Science Resilience	.71	.74	.73	.70	.37	.29	.29	.36	.63	--
D Score	.46	.35	.29	.35	.22	-.01	.15	.18	.23	.40

Table 12. Correlation Matrix for Identification Survey Items and D Score for Women

Item	Science Interest	Science Ability	Confidence in Science Ability	Science Identification	Overall Academic Ability	Confidence in Overall Academic Ability	Satisfaction with Major/Discipline	Persistence in Major	Science Persistence	Science Resilience
Science Ability	.68	--								
Confidence in Science Ability	.63	.86	--							
Science Identification	.75	.79	.72	--						
Overall Academic Ability	.25	.41	.42	.30	--					
Confidence in Overall Academic Ability	.26	.37	.43	.28	.62	--				
Satisfaction with Major/Discipline	.27	.33	.27	.29	.34	.46	--			
Persistence in Major	.28	.33	.26	.23	.45	.37	.61	--		
Science Persistence	.75	.67	.67	.74	.28	.24	.30	.35	--	
Science Resilience	.73	.67	.59	.70	.21	.25	.25	.25	.79	--
D Score	-.22	-.15	-.17	-.16	.06	.05	.03	.11	-.09	-.18

Table 13. Means, Standard Deviations, and Reliability Estimates for Identification Survey Items

Item/Scales	Mean Score (SD)	Reliability (Cronbach's α)	N
Science Identification Items	3.45 (1.02)	.942	307
Science Interest	3.56 (1.19)		310
Science Ability	3.55 (1.08)		310
Confidence in Science Ability	3.29 (1.14)		310
Science Identification	3.39 (1.17)		310
Science Persistence	3.35 (1.13)		307
Science Resilience	3.62 (1.28)		309
General Academics Items	3.97 (.60)	.784	309
Overall Academic Ability	3.95 (.70)		310
Confidence in Overall Academic Ability	3.78 (.73)		310
Satisfaction with Major/Discipline	4.06 (.87)		310
Persistence in Major/Discipline	4.08 (.79)		309
All Items	3.66 (.75)	.904	307

Table 14. Hierarchical Multiple Regression Results for Men

Independent Variable	Final Weights		
	β	t	p
1. Control Variables			
Discipline	-.108	-1.10	.274
Student-Faculty Status	.156	1.68	.097
Experimental Condition	-.028	.32	.752
2. General Academics Variables			
Overall Ability	.155	1.26	.211
Confidence in Overall Ability	-.370	-3.11	.002**
Satisfaction with Major/Discipline	-.049	-.44	.661
Persistence in Major/Discipline	.167	1.37	.174
3. Science Identification Variables			
Science Interest	.393	2.35	.021*
Science Ability	.079	.39	.696
Confidence in Science Ability	-.136	-.71	.481
Identification with Science	.097	.54	.588
Science Persistence	-.255	-1.86	.066
Science Resilience	.285	1.94	.056
<p>Notes: The dependent variable was the D score. The first step was marginally not significant, $F(3, 103) = 2.51$, $p = .063$, $R^2 = .07$. The second step was significant, $F(7, 99) = 2.42$, $p = .025$, $R^2 = .15$, $\Delta R^2 = .08$. The third and final step was also significant, $F(13, 93) = 3.79$, $p < .001$, $R^2 = .35$, $\Delta R^2 = .20$.</p> <p>Column headings: β = beta weight, t = t-test value, p = probability</p> <p>*$p < .05$, **$p < .01$</p>			

Table 15. Hierarchical Multiple Regression Results for Women

Independent Variable	Final Weights		
	β	t	p
1. Control Variables			
Discipline	-.072	-1.00	.317
Student-Faculty Status	.139	1.74	.083
Experimental Condition	.098	-1.37	.172
2. General Academics Variables			
Overall Ability	.021	.22	.830
Confidence in Overall Ability	.113	1.14	.255
Satisfaction with Major/Discipline	-.167	-1.77	.078
Persistence in Major/Discipline	.173	1.74	.083
3. Science Identification Variables			
Science Interest	-.352	-2.93	.004**
Science Ability	.133	.84	.405
Confidence in Science Ability	-.283	-1.94	.054
Identification with Science	.034	.26	.799
Science Persistence	.317	2.22	.028*
Science Resilience	-.154	-1.23	.220
<p>Notes: The dependent variable was the D score. The first step was not significant, $F(3, 189) = 1.55$, $p = .203$, $R^2 = .02$. The second step was also not significant, $F(7, 185) = 1.41$, $p = .205$, $R^2 = .05$, $\Delta R^2 = .03$. The third and final step was significant, $F(13, 179) = 2.51$, $p = .004$, $R^2 = .15$, $\Delta R^2 = .10$. Column headings: β = beta weight, t = t-test value, p = probability *$p < .05$, **$p < .01$</p>

Figure 1. Theoretical Model of Under-representation of Women in STEM Disciplines

Figure 2. Diagram of Anticipated Prediction Model

Figure 3. Means Plot for Gender and Experimental Condition

Figure 4. Means Plot for Discipline and Experimental Condition

Figure 5. Means Plot for Student-Faculty Status and Experimental Condition

Figure 6. Revised Means Plot for Discipline and Experimental Condition

Figure 7. Revised Means Plot for Student-Faculty Status and Experimental Condition

Figure 8. Mediation Model with Measurement and Structural Paths for the Full Sample

Notes: The model chi-square was significant, $\chi^2(85, N=300) = 357.82, p < .001$, CFI = .88, and RMSEA = 0.10. In Figures 8-14, the D score represents implicit endorsement of gender-science stereotypes.

Figure 9. Mediation Model with Measurement and Structural Paths for Men

Note: The model chi-square was significant, $\chi^2(77, N = 107) = 148.68, p < .001$, CFI = .91, and RMSEA = 0.09.

Figure 10. Full Model with Measurement and Structural Paths for Men

Note: The model chi-square was significant, $\chi^2(76, N = 107) = 148.25, p < .001$, CFI = .91, and RMSEA = 0.09.

Figure 11. Direct Effects Model with Measurement and Structural Paths for Men

Note: The model chi-square was significant, $\chi^2(78, N = 107) = 201.72, p < .001$, CFI = .85, and RMSEA = 0.12.

Figure 12. Post-Hoc Adjusted Mediation Model with Measurement and Structural Paths for Men

Note: The model chi-square was significant, $\chi^2(73, N = 107) = 121.69, p < .001$, CFI = .94, and RMSEA = 0.08.

Figure 13. Mediation Model with Measurement and Structural Paths for Women

Note: The model chi-square was significant, $\chi^2(77, N = 193) = 304.60, p < .001$, CFI = .83, and RMSEA = 0.12.

Figure 14. Post-Hoc Adjusted Mediation Model with Measurement and Structural Paths for Women

Note: The model chi-square was significant, $\chi^2(74, N = 193) = 250.57, p < .001$, CFI = .87, and RMSEA = 0.11.

APPENDICES

Appendix A. IAT Instructions for Blocks 1-7

Block 1

In the next task, you will be presented with a set of words or images to classify into groups. Each word or image has a correct classification. Most of these are easy. This will take about 5 minutes. For best results, avoid distractions and stay focused.

Put your middle or index fingers on the E and I keys of your keyboard. Labels at the top will tell you which words go with each key. Keep your index fingers on the 'e' and 'i' keys to enable rapid response.

Words representing the categories at the top will appear one-by-one in the middle of the screen. When the item belongs to a category on the left, press the E key; when the item belongs to a category on the right, press the I key. Items belong to only one category. If you make an error, an X will appear - fix the error by hitting the other key.

This is a timed sorting task. GO AS FAST AS YOU CAN while making as few mistakes as possible. Going too slow or making too many errors will result in an uninterpretable score.

Block 2

See above, the categories have changed. The items for sorting have changed as well. The rules, however, are the same.

When the item belongs to a category on the left, press the E key; when the item belongs to a category on the right, press the I key. Items belong to only one category. An X appears after an error - fix the error by hitting the other key. GO AS FAST AS YOU CAN.

Block 3

See above, the four categories you saw separately now appear together. Remember, each item belongs to only one group. For example, if the categories flower and good appeared on the separate sides above - words meaning flower would go in the flower category, not the good category.

The green and white labels and items may help to identify the appropriate category. Use the E and I keys to categorize items into four groups left and right, and correct errors by hitting the other key.

Block 4

Sort the same four categories again. Remember to go as fast as you can while making as few mistakes as possible.

The green and white labels and items may help to identify the appropriate category. Use the E and I keys to categorize items into the four groups left and right, and correct errors by hitting the other key.

Block 5

Notice above, there are only two categories and they have switched positions. The concept that was previously on the left is now on the right, and the concept that was on the right is now on the left. Practice this new configuration.

Use the E and I keys to categorize items left and right, and correct errors by hitting the other key.

Block 6

See above, the four categories now appear together in a new configuration. Remember, each item belongs to only one group.

The green and white labels and items may help to identify the appropriate category. Use the E and I keys to categorize items into the four groups left and right, and correct errors by hitting the other key.

Block 7

Sort the same four categories again. Remember to go as fast as you can while making as few mistakes as possible.

The green and white labels and items may help to identify the appropriate category. Use the E and I keys to categorize items into the four groups left and right, and correct errors by hitting the other key.

Appendix B. Categories and Items Used in the IAT

Male names: Adam, Chip, Harry, Josh, Roger, Alan, Frank, Ian, Justin, Ryan, Andrew, Fred, Jack, Matthew, Stephen, Brad, Greg, Jed, Paul, Todd, Brandon, Hank, Jonathan, Peter, Wilbur

Female names: Amanda, Courtney, Heather, Melanie, Sara, Amber, Crystal, Katie, Meredith, Shannon, Betsy, Donna, Kristin, Nancy, Stephanie, Bobbi-Sue, Ellen, Lauren, Peggy, Sue-Ellen, Colleen, Emily, Megan, Rachel, Wendy

Science disciplines: Astronomy, Biochemistry, Biology, Chemistry, Engineering, Neuroscience, Physics, Science

Humanities disciplines: Arts, English, History, Humanities, Latin, Music, Philosophy, Spanish

Appendix C. Measures

Demographic Questions

Gender: Male Female

Age: _____

Race/Ethnicity: Asian White Black Hispanic Native American More than One Other

Occupation (select one):

Student

Staff

Faculty

(Please indicate:

Undergraduate Graduate)

adjunct other)

(Please indicate:
full associate assistant

Discipline/major: _____

Identification Survey

1. How would you rate your interest in math and science?

1 = Low interest

2 = Slight interest

3 = Some interest

4 = High interest

5 = Very high interest

2. How would you rate your math and science ability?

1 = Low ability

2 = Slight ability

3 = Some ability

4 = High ability

5 = Very high ability

3. How would you rate your confidence in your math and science ability?

1 = Low confidence

2 = Slight confidence

3 = Some confidence

4 = High confidence

5 = Very high confidence

4. How would you rate your identification with math and science?

1 = Low identification

2 = Slight identification

3 = Some identification

- 4 = High identification
- 5 = Very high identification

5. How would you rate your overall academic or scholarly ability?

- 1 = Low ability
- 2 = Slight ability
- 3 = Some ability
- 4 = High ability
- 5 = Very high ability

6. How would you rate your confidence in overall ability?

- 1 = Low confidence
- 2 = Slight confidence
- 3 = Some confidence
- 4 = High confidence
- 5 = Very high confidence

7. How satisfied are you with your major/discipline?

- 1 = Low satisfaction
- 2 = Slight satisfaction
- 3 = Some satisfaction
- 4 = High satisfaction
- 5 = Very high satisfaction

8. How would you rate your persistence in your major/discipline?

- 1 = Low persistence
- 2 = Slight persistence
- 3 = Some persistence
- 4 = High persistence
- 5 = Very high persistence

9. How would you rate your persistence in math and science?

- 1 = Low persistence
- 2 = Slight persistence
- 3 = Some persistence
- 4 = High persistence
- 5 = Very high persistence

10. How would you rate the likelihood that you will continue in math and science even if you face difficulties?

- 1 = Low likelihood
- 2 = Slight likelihood
- 3 = Some likelihood
- 4 = High likelihood
- 5 = Very high likelihood

Familiarity with Implicit Association Tests survey

Have you completed an Implicit Associations Test before? Yes No

If yes, how many? _____

Do you think this task measured explicit knowledge? Yes No

Do you think this task measured unconscious knowledge? Yes No

Do you think this task measured general knowledge? Yes No

Do you think this task measured something about women scientists? Yes No

Do you think this task measured stereotypes? Yes No

Do you think this task measured something else? (specify in the box) _____

Appendix D. STEM and Non-STEM Disciplines

Notes: The following codes were used by the ADVANCE Program at the University of Rhode Island. Not all departments were represented in the sample used in this study. For some analyses, social and behavioral departments (SBS; Psychology and Sociology & Anthropology) were analyzed separately from STEM departments.

STEM departments include:

- Chemistry
- Computer Science & Statistics
- Mathematics
- Physics
- Psychology (SBS)
- Sociology & Anthropology (SBS)
- Chemical Engineering
- Civil Engineering
- Electrical Engineering
- Industrial Engineering
- Mechanical Engineering
- Ocean Engineering
- Biological Sciences
- Cell & Molecular Biology
- Environmental & Natural Resource Economics
- Fisheries, Animal & Veterinary Science
- Geosciences
- Marine Affairs
- Natural Resource Science
- Plant Science
- Biomedical Sciences
- Oceanography
- Economics
- Education
- English
- English Language Studies
- Film Studies
- Finance & Insurance
- General Business
- Gerontology
- History
- Human Development & Family Studies
- International Business
- Journalism
- Kinesiology
- Labor Relations & Human Resources
- Library & Information Studies
- Management
- Management Information System
- Marketing
- Military Science
- Modern & Classical Languages
- Music
- Nursing
- Nutrition & Food Sciences
- Pharmacy Practice
- Philosophy
- Physical Education & Exercise Science
- Physical Therapy
- Political Science
- Textile, Fashion Merchandising & Design
- Theater
- Women's Studies
- Writing

Non-STEM departments include:

- Accounting
- African & Afro-American Studies
- Art
- Clinical Laboratory Science
- Communication Studies
- Communicative Disorders
- Community Planning & Landscape Architecture
- Dental Hygiene

Appendix E. Informed Consent Form

Implicit Associations and Majors/Disciplines: Informed Consent

Dear Participant,

You have been asked to take part in the project described below. You were selected as a possible participant because you are a student or faculty member at the University of Rhode Island. Please read this form and ask any questions that you may have before agreeing to be in the research.

The purpose of the study is to examine implicit, or unconscious associations and majors/disciplines. The study will measure reaction times on a categorization task. The study will also examine attitudes about majors/disciplines and self-esteem.

Karen Stamm, the student investigator, and Dr. Lisa Harlow, the faculty sponsor, at the University of Rhode Island are conducting this study.

YOU MUST BE AT LEAST 18 YEARS OLD to participate this research project.

If you decide to take part in this study, you will be asked to do the following things:

- You will be asked to complete a web-based computerized categorization task. Words will be shown to you on the computer screen. Each word belongs to one of two categories (i.e., flowers or insects). As quickly as possible, you will be asked to classify each word.
- You will also be asked to answer some brief questionnaires about yourself.

The study will take approximately 15 minutes or less. After your participation today, you will not be asked to return for further participation.

The possible risks or discomforts of the study are minimal.

Although there are no direct benefits of the study, your answers will enable us to provide information on attitudes about majors/disciplines.

Your participation in this study is anonymous and confidential. That means that your answers to all questions are private. No one else can know if you participated in this study and no one else can find out what your answers were. You will not be asked to provide your name, phone number, email address, or other identifying information. There will be no link between data and participant identity. Data will be collected by a web-based survey and research study program designed by Millisecond Software. Once downloaded for analysis, data will be password protected and stored on Karen Stamm's personal computer. Scientific reports will be based on group data and will not identify you or any individual as being in this project. All information in the surveys will be strictly confidential. Data will be summarized and carefully reported so that no single individual can be identified.

The decision to participate in this research project is up to you. You do not have to participate and you can refuse to answer any question.

Participation in this study is not expected to be harmful or injurious to you. However, if this study causes you any injury, or if you have any questions, you should contact Karen Stamm at stammka@mail.uri.edu. You may also contact Dr. Lisa Harlow at (401) 874-4242 or lharlow@uri.edu.

If you have other concerns about this study, you may contact the University of Rhode Island's Vice President for Research, 70 Lower College Road, Suite 2, URI, Kingston, RI, (401) 874-4328.

You have read the consent form and your questions have been answered to your satisfaction. Your participation in the categorization task and your filling out the survey implies your consent to participate in this study.

Thank you for your participation.

*Participate in a web-based research
study on implicit associations
and majors/disciplines*

- You must be at least 18 years old.
- You must be a student or faculty member at the University of Rhode Island.
- Your responses are anonymous and confidential. You will not be asked to provide your name or other identifying information.
- Your participation will involve responding as quickly as possible to a categorization task and answering some questions about yourself.
- Participation will take approximately 10-15 minutes.

To participate in this study, go to:
<http://research.millisecond.com/stammka/iat4.web>

Please contact Karen Stamm (stammka@mail.uri.edu) with any questions.

Thank you for your participation!

BIBLIOGRAPHY

- ADVANCE Institutional Transformation Indicators Working Group. (2005). Proposed toolkit for reporting progress toward NSF ADVANCE: Institutional transformation goals. Retrieved June 14, 2009 from http://www.advance.vt.edu/Measuring_Progress/Toolkits/Advance_Data_Indicators_Toolkit_Jan05.pdf
- Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40*, 401–408.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat in African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113-125.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and Company.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136, 256-276.
- Bem, S. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42, 155-162.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows user's guide*. Encino, CA: Multivariate Software.
- Betz, N. E., & Hackett, G. (1981). The relationship of career-related self efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology*, 28, 399-410.
- Betz, N. E., & Schifano, R. S. (2000). Evaluation of an intervention to increase realistic self-efficacy and interests in college women. *Journal of Vocational Behavior*, 56, 35-52.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, 81, 828-841.
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity* (pp.35-58). Oxford, United Kingdom: Blackwell.
- Brown, S. D., Tramayne, S., Hoxha, D., Telander, K., Fan, X., & Lent, R. W. (2008). Social cognitive predictors of college students' academic performance and

- persistence: A meta-analytic path analysis. *Journal of Vocational Behavior*, 72, 298–308.
- Burkley, M., & Blanton, H. (2008). Endorsing a negative in-group stereotype as a self-protective strategy: Sacrificing the group to save the self. *Journal of Experimental Social Psychology*, 44, 37-49.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135, 218-265.
- Cohen, J., Cohen, P., West, S. W., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cokley, K. O. (2002). Ethnicity, gender, and academic self-concept: A preliminary examination of academic disidentification and implications for psychologists. *Cultural Diversity and Ethnic Minority Psychology*, 8, 378-388.
- Commission on Professionals in Science and Technology. (2007). *Four decades of STEM degrees, 1966-2004: "The devil is in the details"* (STEM Workforce Data Project. No. 6). Retrieved November 16, 2007 from https://www.cpst.org/STEM/STEM6_Report.pdf.
- Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavior Research*, 33, 295-312.
- Crisp, R. J., & Turner, R. N. (2009). Can imagined interactions produce positive perceptions? Reducing prejudice through simulated social contact. *American Psychologist*, 64, 231-240.

- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220-230.
- Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance, 19*, 421-440.
- Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3*, 286-300.
- Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology, 93*, 764-779.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642-658.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615-1628.
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88*, 276-287.

- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Dietz-Uhler, B., & Murrell, A. (1998). Effects of social identity and threat on self-esteem and group attributions. *Group Dynamics: Theory, Research, and Practice*, 2, 24-35.
- Dominici, F., Fried, L. P., & Zeger, S. L. (2009). So few women leaders. *Academe*, 95(4), 25-27.
- Donovan, C., Hodgson, B., Scanlon, E., & Whitelegg, E. (2005). Women in higher education: Issues and challenges for part-time scientists. *Women's Studies International Forum*, 28, 247-258.
- Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. *Cultural Diversity and Ethnic Minority Psychology*, 8, 88-102.
- Dovidio, J. F., Gaertner, S. L., & Validzic, A. (1998). Intergroup bias: Status, differentiation, and a common in-group identity. *Journal of Personality and Social Psychology*, 75, 109-120.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality Psychology*, 82, 62-68.
- Drago, R., Colbeck, C. L., Stauffer, K. D., Pirretti, A., Burkum, K., Fazioli, J., Lazzaro, G., & Habasevich, T. (2006). The avoidance of bias against caregiving: The case of academic faculty. *American Behavioral Scientist*, 49, 1222-1247.

- Feist, G. J. (2006). How development and personality influence scientific thought, interest, and achievement. *Review of General Psychology. Special Issue: The Psychology of Science, 10*, 163-182.
- Fiske, S. T. (2004). *Social beings: A core motives approach to social psychology*. Hoboken, NJ: Wiley.
- Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 30*, 1611-1624.
- Hrabowski, III, F. A., & Maton, K. I. (2009). Change institutional culture, and you change who goes into science. *Academe, 95*(3), 11-15.
- Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17-28.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93-104.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3-25.

- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022-1038.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- Gresky, D. M., Ten Eyck, L. L., Lord, C. G., & McIntyre, R. B. (2005). Effects of salient multiple identities on women's performance under mathematics stereotype threat. *Sex Roles*, 53, 703-716.
- Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, 92, 81-92.
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89, 416-427.
- Herzig, A. H. (2006). How can women and students of color come to belong in graduate mathematics? In Bystudzienski, J. M., & Bird, S. R (Eds.), *Removing barriers: Women in academic science, technology, engineering, and mathematics* (pp. 254-270). Bloomington, IN: Indiana University Press.

- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369-1385.
- Hogg, M. A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes*. New York: Routledge.
- Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, 99, 545-560.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, 14, 299-324.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008, July 25). Gender similarities characterize math performance. *Science*, 321, 494-495.
- IAT Corp. (2007, March). *Project Implicit*. Retrieved March 31, 2007, from <https://implicit.harvard.edu/implicit/>
- Inquisit 3.0.2.0. [Computer software]. (2008). Seattle, WA: Millisecond Software.
- Jordan, A. H., & Lovett, B. J. (2007). Stereotype threat and test performance: A primer for school psychologists. *Journal of School Psychology*, 45, 45-59.

- Kawakami, K., Steele, J. R., Cifa, C., Phillips, C. E., & Dovidio, J. F. (2008). Approaching math increase math = me and math = pleasant. *Journal of Experimental Social Psychology*, 44, 818-825.
- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting effect on women's math performance. *Personality and Social Psychology Bulletin*, 29, 371-381.
- Kelley, R. (2009, July 13). The roots of racism: What we don't know can hurt us. *Newsweek*, 154(2), 27.
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental and Social Psychology*, 43, 825-832.
- Kite, M. E., Russo, N. F., Brehm, S. S., Fouad, N. A., Hall, C. C. I., Hyde, J. S., & Keita, G. P. (2001). Women psychologists in academe: Mixed progress, unwarranted complacency. *American Psychologist*, 56, 1080-1098.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: What we know (so far) about the method. In B. Wittenbrink & N. Schwartz (Eds.), *Implicit measures of attitudes* (pp. 59-102). New York: Guilford Press.

- Lee, J. D. (1998). Which kids can “become” scientists? Effects of gender, self-concepts and perceptions of scientists. *Social Psychology Quarterly*, 61, 199-219.
- Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math identified women respond to stereotype threat. *Sex Roles*, 54, 113-125.
- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology*, 96, 1089-1103.
- Maccoby, E. M., & Jacklin, C. N. (1974). *The psychology of sex differences (vol. 1)*. Stanford, CA: Stanford University Press.
- Martin, D. W. (2008). *Doing psychology experiments* (7th ed.). Belmont, CA: Thompson Wadsworth.
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women’s math test performance. *Personality and Social Psychology Bulletin*, 28, 1183-1193.
- Marx, D. M., & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, 91, 243-254.
- Morgeson, F. P., Seligman, M. E. P., Sternberg, R. J., Taylor, S. E., & Manning, C. M. (1999). Lessons learned from a life in psychological science: Implications for young scientists. *American Psychologist*, 54, 106-116.
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18, 879-885.

- National Science Foundation (2005a). ADVANCE: Increasing the participation and advancement of women in academic science and engineering careers. Retrieved March 31, 2007, from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5383&from=fund
- National Science Foundation (2005b). *More women receive Ph.D.s, but female senior faculty are still rare*. Retrieved March 31, 2007, from http://www.nsf.gov/news/news_summ.jsp?cntn_id=104363.
- Nguyen, H.-H. D., & Ryan, A. M. (2009). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93, 1314–1334.
- Noordenbos, G. (2002). Women in academies of sciences: From exclusion to exception. *Women's Studies International Forum*, 25, 127-137.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134, 565-584.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice. Special Issue: Groups and Internet*, 6, 101-115.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = male, me = female, therefore math $\neq$ me. *Journal of Personality and Social Psychology*, 83, 44-59.
- Nosek, B. A., Greenwald, A. G. & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166-180.

- Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, 54, 14-29.
- Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology*, 43, 127-134.
- Osbourne, J. W. (1995). Academics, self-esteem, and race: A look at the assumptions underlying the disidentification hypothesis. *Personality and Social Psychology Bulletin*, 21, 449-455.
- Osbourne, J. W. (2006). Gender, stereotype threat, and anxiety: Psychophysiological and cognitive evidence. *Journal of Research in Educational Psychology*, 8, 108-109.
- Osbourne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology*, 27, 135-154.
- Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37, 239-253.
- Pfafflin, S. M. (1984). Women, science, and technology. *American Psychologist*, 39, 1183-1186.
- Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57, 55-71.

- Rice, R. E., Sorcinelli, M. D., & Austin, A. E. (2000). Heeding new voices: Academic careers for a new generation. Retrieved May 23, 2008 from http://eric.ed.gov/ericdocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/c9/b0.pdf
- Rosenbloom, J. L., Ash, R. A., Dupont, B., & Coder, L. (2008). Why are there so few women in information technology? Assessing the role of personality in career choices. *Journal of Economic Psychology*, 29, 543-554.
- Rosser, S. V., & Taylor, M. Z. (2009). Why are we still worried about women in science? *Academe*, 95(3), 6-10.
- Rudd, E., Morrison, E., Picciano, J., & Nerad, M. (2008). Finally equal footing for women in social science careers? *CIRGE Spotlight on Doctoral Education #1*. CIRGE: University of Washington, Seattle, WA. Retrieved August 22, 2008, from <http://depts.washington.edu/cirgeweb/c/research/social-science-phds%E2%80%94five-years-out/>
- Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). “Unlearning” automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81, 856-868.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women’s math performance. *Journal of Experimental Social Psychology*, 38, 194-201.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336-356.

- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology, 39*, 68-74.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*, 70-83.
- Sinclair, S., Hardin, C. D., & Lowery, B. S. (2006). Self-stereotyping in the context of multiple social identities. *Journal of Personality and Social Psychology, 90*, 529-542.
- Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99*, 99-114.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *American Psychologist, 60*, 950-958.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*, 1191-1197.

- Steele, C. M. (1992, April). Race and the schooling of black Americans. *Atlantic Monthly*, 269(4), 68-78.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Steele, J., James, J. B., & Barnett, R. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly*, 26, 46-50.
- Steiger, J. H., & Lind, J. M. (1980, June). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77, 1213-1227.
- Tabachnick, D. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed). New York: Harper Collins.
- Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. New York: Cambridge University Press.
- Thoman, D. B., White, P. H., Yamawaki, N., & Koishi, H. (2008). Variations of gender-math stereotype content affect women's vulnerability to stereotype threat. *Sex Roles*, 58, 702-712.

- University of Rhode Island, Institutional Research. (2008). Just the facts sheet: Fall 2008. Retrieved May 27, 2009 from <http://www.uri.edu/ir/uriinfobank/facts.html>
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.
- Valian, V. (2006). Beyond gender schemas: Improving the advancement of women in academia. In Bystudzienski, J. M., & Bird, S. R (Eds.), *Removing barriers: Women in academic science, technology, engineering, and mathematics* (pp. 320-332). Bloomington, IN: Indiana University Press.
- van Anders, S. M. (2004). Why the academic pipeline leaks: Fewer men than women perceive barriers to becoming professors. *Sex Roles, 51*, 511-521.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826.